(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42436 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/42995

(22) International Filing Date:
7 November 2001 (07.11.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/715,177 20 November 2000 (20.11.2000) US
09/761,640 18 January 2001 (18.01.2001) US

(71) Applicant: **PE CORPORATION (NY)** [US/US]; 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: **WEI, Ming-Hui**; c/o Celera Genomics, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US). **KETCHUM, Karen, A.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **DI FRANCESCO, Valentina**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US). **BEASLEY, Ellen, M.**; c/o Celera Genomics, 45 West Gude Drive C2-4#21, Rockville, MD 20850 (US).

(74) Agent: **MILLMAN, Robert, A.**; CELERA GENOMICS, Chief Intellectual Property Counsel, 45 West Gude Drive C2-4#20, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/42436 A2

(54) Title: ISOLATED HUMAN PHOSPHATASE PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS, AND USES THEREOF

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the phosphatase peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the phosphatase peptides, and methods of identifying modulators of the phosphatase peptides.

ISOLATED HUMAN PHOSPHATASE PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS, AND USES THEREOF

RELATED APPLICATIONS

The present application is a Continuation-In-Part of U.S. Serial No. 09/715,177, filed November 20, 2000 and Continuation-In-Part of U.S. Serial No. 09/761,640.

FIELD OF THE INVENTION

The present invention is in the field of phosphatase proteins that are related to the mitogen-activated protein (MAP) kinase phosphatase subfamily, recombinant DNA molecules and protein production. The present invention provides novel phosphatase peptides and proteins and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods. Specifically, the present invention provides three novel MAP kinase phosphatase splice forms.

BACKGROUND OF THE INVENTION

Phosphatase proteins, particularly members of the MAP kinase phosphatase subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of phosphatase proteins. The present invention advances the state of the art by providing a previously unidentified human phosphatase proteins that have homology to members of the MAP kinase phosphatase subfamily.

Protein Phosphatases

Cellular signal transduction is a fundamental mechanism whereby external stimuli that regulate diverse cellular processes are relayed to the interior of cells. The biochemical pathways through which signals are transmitted within cells comprise a circuitry of directly or functionally connected interactive proteins. One of the key biochemical mechanisms of signal transduction involves the reversible phosphorylation of certain residues on proteins. The phosphorylation state of a protein may affect its conformation and/or enzymic activity as well as its cellular location. The phosphorylation state of a protein is modified through the reciprocal actions of protein phosphatases (PKs) and protein phosphatases (PPs) at various specific amino acid

residues.

Protein phosphorylation is the ubiquitous strategy used to control the activities of eukaryotic cells. It is estimated that 10% of the proteins active in a typical mammalian cell are phosphorylated. The high-energy phosphate that confers activation and is transferred from adenosine triphosphate molecules to protein-by-protein phosphatases is subsequently removed from the protein-by-protein phosphatases. In this way, the phosphatases control most cellular signaling events that regulate cell growth and differentiation, cell-to-cell contacts, the cell cycle, and oncogenesis.

The protein phosphorylation/dephosphorylation cycle is one of the major regulatory mechanisms employed by eukaryotic cells to control cellular activities. It is estimated that more than 10% of the active proteins in a typical mammalian cell are phosphorylated. During protein phosphorylation/dephosphorylation, phosphate groups are transferred from adenosine triphosphate molecules to protein-by-protein phosphatases and are removed from the protein-by-protein phosphatases.

Protein phosphatases function in cellular signaling events that regulate cell growth and differentiation, cell-to-cell contacts, the cell cycle, and oncogenesis. Three protein phosphatase families have been identified as evolutionarily distinct. These include the serine/threonine phosphatases, the protein tyrosine phosphatases, and the acid/alkaline phosphatases (Carbonneau H. and Tonks N. K. (1992) Annu. Rev. Cell Biol. 8:463-93).

The serine/threonine phosphatases are either cytosolic or associated with a receptor. On the basis of their sensitivity to two thermostable proteins, inhibitors 1 and 2, and their divalent cation requirements, the serine/threonine phosphatases can be separated into four distinct groups, PP-I, PP-IIA, PP-IIB, and PP-IIC.

PP-I dephosphorylates many of the proteins phosphorylated by cylic AMP-dependent protein phosphatase and is therefore an important regulator of many cyclic AMP mediated, hormone responses in cells. PP-IIA has broad specificity for control of cell cycle, growth and proliferation, and DNA replication and is the main phosphatase responsible for reversing the phosphorylations of serine/threonine phosphatases. PP-IIB, or calcineurin (Cn), is a Ca.sup.+2 -activated phosphatase; it is involved in the regulation of such diverse cellular functions as ion channel regulation, neuronal transmission, gene transcription, muscle glycogen metabolism, and lymphocyte activation.

PP-IIC is a Mg.sup.++ -dependent phosphatase which participates in a wide variety of functions including regulating cyclic AMP-activated protein-phosphatase activity, Ca.sup.++ -

dependent signal transduction, tRNA splicing, and signal transmission related to heat shock responses. PP-IIC is a monomeric protein with a molecular mass of about 40-45 kDa. One .alpha. and several .beta. isoforms of PP-IIC have been identified (Wenk, J. et al. (1992) FEBS Lett. 297: 135-138; Terasawa, T. et al. (1993) Arch. Biochem. Biophys. 307: 342-349; and Kato, S. et al. (1995) Arch. Biochem. Biophys. 318: 387-393).

The levels of protein phosphorylation required for normal cell growth and differentiation at any time are achieved through the coordinated action of PKs and PPS. Depending on the cellular context, these two types of enzymes may either antagonize or cooperate with each other during signal transduction. An imbalance between these enzymes may impair normal cell functions leading to metabolic disorders and cellular transformation.

For example, insulin binding to the insulin receptor, which is a PTK, triggers a variety of metabolic and growth promoting effects such as glucose transport, biosynthesis of glycogen and fats, DNA synthesis, cell division and differentiation. Diabetes mellitus, which is characterized by insufficient or a lack of insulin signal transduction, can be caused by any abnormality at any step along the insulin signaling pathway. (Olefsky, 1988, in "Cecil Textbook of Medicine," 18th Ed., 2:1360-81).

It is also well known, for example, that the overexpression of PTKs, such as HER2, can play a decisive role in the development of cancer (Slamon et al., 1987, Science 235:77-82) and that antibodies capable of blocking the activity of this enzyme can abrogate tumor growth (Drebin et al., 1988, Oncogene 2:387-394). Blocking the signal transduction capability of tyrosine phosphatases such as Flk-1 and the PDGF receptor have been shown to block tumor growth in animal models (Millauer et al., 1994, Nature 367:577; Ueno et al., Science, 252:844-848).

Relatively less is known with respect to the direct role of phosphatases in signal transduction; PPs may play a role in human diseases. For example, ectopic expression of RPTP.alpha. produces a transformed phenotype in embryonic fibroblasts (Zheng et al., Nature 359:336-339), and overexpression of RPTP.alpha. in embryonal carcinoma cells causes the cells to differentiate into a cell type with neuronal phenotype (den Hertog et al., EMBO J 12:3789-3798). The gene for human RPTP.gamma. has been localized to chromosome 3p21 which is a segment frequently altered in renal and small lung carcinoma. Mutations may occur in the extracellular segment of RPTP.gamma. which renders a RPTP that no longer respond to external signals (LaForgia et al., Wary et al., 1993, Cancer Res 52:478-482). Mutations in the gene encoding PTP1C (also known as HCP, SHP) are the cause of the moth-eaten phenotype in mice

that suffer severe immunodeficiency, and systemic autoimmune disease accompanied by hyperproliferation of macrophages (Schultz et al., 1993, Cell 73:1445-1454). PTP1D (also known as Syp or PTP2C) has been shown to bind through SH2 domains to sites of phosphorylation in PDGFR, EGFR and insulin receptor substrate 1 (IRS-1). Reducing the activity of PTP1D by microinjection of anti-PTP1D antibody has been shown to block insulin or EGF-induced mitogenesis (Xiao et al., 1994, J Biol Chem 269:21244-21248).

MAP Kinase Phosphatases

The present invention provides three novel alternative splice forms of mitogen-activated protein (MAP) kinase phosphatase. The alternative splice forms are herein referred to as splice forms 1, 2, and 3. Specifically, as indicated in Figure 3, splice form 2 includes exons 13 and 14, which are absent in splice forms 1 and 3, and splice form 3 is missing exon 7, which is present in splice forms 1 and 2. cDNA clones from all three isoforms have been mapped to the same region of human chromsome 11. Splice form 1 has been previously disclosed by applicant in U.S. application 09/715,177, filed November 20, 2000.

MAP kinase phosphatases are dual-specificity protein phosphatases involved in numerous critical biological processes. In *Drosophila*, MAP kinase phosphatases have been found to be essential for viability. Furthermore, loss-of-function mutations cause kinked and/or branched bristles and wing hairs. In plants, MAP kinase phosphatases play a critical role in responses to stress and pathogens (Gupta *et al., Plant J* 1998 Dec;16(5):581-9). MAP kinase phosphatases play a critical role in neuronal survival and neuronal cell death following injury and degenerative stimuli (Winter *et al., Brain Res* 1998 Aug 10;801(1-2):198-205). MAP kinase phosphatase may also play a role in diabetes and other insulin-related conditions; insulin regulates map kinase phosphatase-1 (MKP-1) and it has been suggested that MKP-1 acts as a negative regulator of insulin signaling (Kusari *et al., Mol Endocrinol* 1997 Sep;11(10):1532-43). Furthermore, MAP kinase phosphatases may be critical for the negative regulation of cellular proliferation (Emslie *et al., Hum Genet* 1994 May;93(5):513-6) and therefore, novel human MAP kinase phosphatase genes are useful as candidate tumor-suppressor genes. Additionally, MAP kinase phosphatases may specifically be involved in pancreatic cancer (Furukawa *et al., Cytogenet Cell Genet* 1998;82(3-4):156-9).

For a further review of MAP kinase phosphatases, see Scimeca *et al., Oncogene* 1997 Aug 7;15(6):717-25. See Wang *et al., Genomics* 57 (2), 310-315 (1999) for supporting information on alternative splice variants.

The discovery of new human protein phosphatases and the polynucleotides encoding them satisfies a need in the art by providing new compositions that are useful in the diagnosis, prevention and treatment of biological processes associated with abnormal or unwanted protein phosphorylation.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human phosphatase peptides and proteins that are related to the MAP kinase phosphatase subfamily, as well as allelic variants and other mammalian orthologs thereof. Specifically, the present invention provides three novel MAP kinase phosphatase splice forms. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate phosphatase activity in cells and tissues that express the phosphatase. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of three cDNA molecules that encode each of the phosphatase splice forms of the present invention. (splice form 1 = SEQ ID NO:1, splice form 2 = SEQ ID NO:4, splice form 3 = SEQ ID NO:6) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain.

FIGURE 2 provides the predicted amino acid sequence of the three phosphatase splice forms of the present invention. (splice form 1 = SEQ ID NO:2, splice form 2 = SEQ ID NO:5, splice form 3 = SEQ ID NO:7) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the three phosphatase splice forms of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. Figure 3 also indicates that the map position of the genomic sequence encoding the three splice forms is human chromosome 11. As illustrated in Figure 3, the following SNP variations were identified: G577A, G1451A, and G2641A. Figure 3 further provides a gene structure model and multiple alignments of the cDNA and peptide sequences of the three splice forms to illustrate the structure/sequence variations. Specifically, as indicated in Figure 3, splice form 2 includes exons 13 and 14, which are absent in splice forms 1 and 3, and splice form 3 is missing exon 7, which is present in splice forms 1 and 2.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a phosphatase protein or part of a phosphatase protein and are related to the MAP kinase phosphatase subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of three novel human MAP kinase phosphatase splice forms, nucleic acid sequences in the form of cDNA sequences and genomic sequences that encode these phosphatase splice forms, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the phosphatase proteins of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known phosphatase proteins of the MAP kinase phosphatase subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known MAP kinase phosphatase family or subfamily of phosphatase proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the phosphatase family of proteins and are related to the MAP kinase phosphatase subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). Specifically, the present invention provides three novel MAP kinase phosphatase splice forms. The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the phosphatase peptides of the present invention, phosphatase peptides, phosphatase splice forms, or peptides/proteins/splice forms of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the phosphatase peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these

peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the phosphatase peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated phosphatase peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. For example, a nucleic acid molecule encoding the phosphatase peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2,5,7), for example, proteins encoded by the cDNA

nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1,4,6) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2,5,7), for example, proteins encoded by the cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1,4,6) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NOS:2,5,7), for example, proteins encoded by the cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1,4,6) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the phosphatase peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The phosphatase peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a phosphatase peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the phosphatase peptide. "Operatively linked" indicates that the phosphatase peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the phosphatase peptide.

In some uses, the fusion protein does not affect the activity of the phosphatase peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant phosphatase peptide. In certain host cells (e.g., mammalian host cells),

expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A phosphatase peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the phosphatase peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the phosphatase peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide

as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength =

3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the phosphatase peptides of the present invention as well as being encoded by the same genetic locus as the phosphatase peptide provided herein. The gene encoding the novel phosphatase proteins of the present invention is located on public BAC AP001885, which is known to be mapped to chromosome 11 (as indicated in Figure 3).

Allelic variants of a phosphatase peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the phosphatase peptide as well as being encoded by the same genetic locus as the phosphatase peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. The gene encoding the novel phosphatase proteins of the present invention is located on public BAC AP001885, which is known to be mapped to chromosome 11 (as indicated in Figure 3). As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a phosphatase peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the phosphatase proteins of the present invention. The following SNPs were identified: G577A, G1451A, and G2641A. G577A and G1451A are non-synonymous coding SNPs. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. G2641A is 3' of the ORF and may affect control/regulatory elements.

Paralogs of a phosphatase peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the phosphatase peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will

typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a phosphatase peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a phosphatase peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the phosphatase peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a phosphatase peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the phosphatase peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the phosphatase peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a phosphatase peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant phosphatase peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to dephosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as phosphatase activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the phosphatase peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a phosphatase peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the phosphatase peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the phosphatase peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in phosphatase peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-

ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the phosphatase peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature phosphatase peptide is fused with another compound, such as a compound to increase the half-life of the phosphatase peptide, or in which the additional amino acids are fused to the mature phosphatase peptide, such as a leader or secretory sequence or a sequence for purification of the mature phosphatase peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as,

for example, in a phosphatase-effector protein interaction or phosphatase-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, phosphatases isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the phosphatase. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. A large percentage of pharmaceutical agents are being developed that modulate the activity of phosphatase proteins, particularly members of the MAP kinase phosphatase subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is

expressed in fetal brain. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to phosphatases that are related to members of the MAP kinase phosphatase subfamily. Such assays involve any of the known phosphatase functions or activities or properties useful for diagnosis and treatment of phosphatase-related conditions that are specific for the subfamily of phosphatases that the one of the present invention belongs to, particularly in cells and tissues that express the phosphatase. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the phosphatase, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the phosphatase protein.

The polypeptides can be used to identify compounds that modulate phosphatase activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the phosphatase. Both the phosphatases of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the phosphatase. These compounds can be further screened against a functional phosphatase to determine the effect of the compound on the phosphatase activity.

Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the phosphatase to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the phosphatase protein and a molecule that normally interacts with the phosphatase protein, e.g. a substrate or a component of the signal pathway that the phosphatase protein normally interacts (for example, another phosphatase). Such assays typically include the steps of combining the phosphatase protein with a candidate compound under conditions that allow the phosphatase protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the phosphatase protein and the target, such as any of the associated effects of signal transduction such as protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant phosphatases or appropriate fragments containing mutations that affect phosphatase function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) phosphatase activity. The assays typically involve an assay of events in the signal transduction pathway that indicate phosphatase activity. Thus, the dephosphorylation of a substrate, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the phosphatase protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the phosphatase can be used as

an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the phosphatase can be assayed. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid.

Binding and/or activating compounds can also be screened by using chimeric phosphatase proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate then that which is recognized by the native phosphatase. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the phosphatase is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the phosphatase (e.g. binding partners and/or ligands). Thus, a compound is exposed to a phosphatase polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble phosphatase polypeptide is also added to the mixture. If the test compound interacts with the soluble phosphatase polypeptide, it decreases the amount of complex formed or activity from the phosphatase target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the phosphatase. Thus, the soluble polypeptide that competes with the target phosphatase region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the phosphatase protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of phosphatase-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a phosphatase-binding protein and a candidate compound are incubated in the phosphatase protein-presenting wells and the amount of complex trapped in the well can be quantitated.. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the phosphatase protein target molecule, or which are reactive with phosphatase protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the phosphatases of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of phosphatase protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the kinase pathway, by treating cells or tissues that express the phosphatase. Experimental data as provided in Figure 1 indicates

that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. These methods of treatment include the steps of administering a modulator of phosphatase activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the phosphatase proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the phosphatase and are involved in phosphatase activity. Such phosphatase-binding proteins are also likely to be involved in the propagation of signals by the phosphatase proteins or phosphatase targets as, for example, downstream elements of a kinase-mediated signaling pathway. Alternatively, such phosphatase-binding proteins are likely to be phosphatase inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a phosphatase protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a phosphatase-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the phosphatase protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified

as described herein (e.g., a phosphatase-modulating agent, an antisense phosphatase nucleic acid molecule, a phosphatase-specific antibody, or a phosphatase-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The phosphatase proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. The method involves contacting a biological sample with a compound capable of interacting with the phosphatase protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered phosphatase activity in cell-based or cell-free assay, alteration in substrate or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the phosphatase protein in which one or more of the phosphatase functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are more or less active in substrate binding, and phosphatase activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1

indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Accordingly, methods for treatment include the use of the phosphatase protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the phosphatase proteins. Antibodies can be prepared from any region of the peptide as described herein.

However, preferred regions will include those involved in function/activity and/or phosphatase/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot

shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment

modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the phosphatase peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nuleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a phosphatase peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the phosphatase peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 [SEQ ID NOS:1,4,6 (cDNA sequences) and SEQ ID NO:3 (genomic sequence)], or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2,5,7. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 [SEQ ID NOS:1,4,6 (cDNA sequences) and SEQ ID NO:3 (genomic sequence)], or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2,5,7. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide

sequences shown in Figure 1 or 3 [SEQ ID NOS:1,4,6 (cDNA sequences) and SEQ ID NO:3 (genomic sequence)], or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2,5,7. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the phosphatase peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of

mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the phosphatase proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that

hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. The gene encoding the novel phosphatase proteins of the present invention is located on public BAC AP001885, which is known to be mapped to chromosome 11 (as indicated in Figure 3).

Figure 3 provides information on SNPs that have been found in the gene encoding the phosphatase proteins of the present invention. The following SNPs were identified: G577A, G1451A, and G2641A. G577A and G1451A are non-synonymous coding SNPs. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. G2641A is 3' of the ORF and may affect control/regulatory elements.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization

probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, the following SNP variations were identified: G577A, G1451A, and G2641A.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. The gene encoding the novel phosphatase proteins of the present invention is located on public BAC AP001885, which is known to be mapped to chromosome 11 (as indicated in Figure 3).

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in phosphatase protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a phosphatase protein, such as by measuring a level of a phosphatase-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a phosphatase gene has been mutated. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid.

Nucleic acid expression assays are useful for drug screening to identify compounds that

modulate phosphatase nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the phosphatase gene, particularly biological and pathological processes that are mediated by the phosphatase in cells and tissues that express it. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. The method typically includes assaying the ability of the compound to modulate the expression of the phosphatase nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired phosphatase nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the phosphatase nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for phosphatase nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the phosphatase protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of phosphatase gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of phosphatase mRNA in the presence of the candidate compound is compared to the level of expression of phosphatase mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate phosphatase nucleic acid expression in cells and tissues that express the phosphatase. Experimental data as provided in

Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for phosphatase nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the phosphatase nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates that splice form 1 of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the phosphatase gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in phosphatase nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in phosphatase genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization

probes to detect naturally occurring genetic mutations in the phosphatase gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the phosphatase gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a phosphatase protein.

Individuals carrying mutations in the phosphatase gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the phosphatase proteins of the present invention. The following SNPs were identified: G577A, G1451A, and G2641A. G577A and G1451A are non-synonymous coding SNPs. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. G2641A is 3' of the ORF and may affect control/regulatory elements. The gene encoding the novel phosphatase proteins of the present invention is located on public BAC AP001885, which is known to be mapped to chromosome 11 (as indicated in Figure 3). Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a phosphatase gene can be directly identified, for example, by

alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant phosphatase gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl. 9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature 313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the phosphatase gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the phosphatase proteins of the present invention. The following SNPs were identified: G577A, G1451A, and G2641A. G577A and G1451A are non-synonymous coding SNPs. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a

reference. G2641A is 3' of the ORF and may affect control/regulatory elements.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control phosphatase gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of phosphatase protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into phosphatase protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of phosphatase nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired phosphatase nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the phosphatase protein, such as substrate binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in phosphatase gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired phosphatase protein to treat the individual.

The invention also encompasses kits for detecting the presence of a phosphatase nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that splice form 1 of the phosphatase protein of the present invention is expressed in humans in the pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, ovary tumor tissue, brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. Additionally, splice form 3 is expressed in fetal brain. Specifically, a virtual northern blot shows expression in pancreas, colon and pancreas adenocarcinomas, pancreas epitheliod carcinomas, lung large cell carcinomas, renal cell carcinomas, placenta choriocarcinomas, and ovary tumor tissue. In addition, PCR-based tissue screening panels indicate expression in the brain (including fetal), heart (including fetal), kidney (including fetal), uterus, and thyroid. For example, the kit can comprise reagents such as a labeled

or labelable nucleic acid or agent capable of detecting phosphatase nucleic acid in a biological sample; means for determining the amount of phosphatase nucleic acid in the sample; and means for comparing the amount of phosphatase nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect phosphatase protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1,3,4,6).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are

unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct

sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the phosphatase proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the phosphatase gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the phosphatase proteins of the present invention. The following SNPs were identified: G577A, G1451A, and G2641A. G577A and G1451A are non-synonymous coding SNPs. Changes in the amino acid sequence caused by these SNPs is indicated in Figure 3 and can readily be determined using the universal genetic code and the protein sequence provided in Figure 2 as a reference. G2641A is 3' of the ORF and may affect control/regulatory elements.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close

confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified phosphatase gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate

cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterophosphatase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant

protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as phosphatases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with phosphatases, the protein can be isolated from the host cell by standard disruption procedures,

including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a phosphatase protein or peptide that can be further purified to produce desired amounts of phosphatase protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the phosphatase protein or phosphatase protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native phosphatase protein is useful for assaying compounds that stimulate or inhibit phosphatase protein function.

Host cells are also useful for identifying phosphatase protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant phosphatase protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native phosphatase protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a phosphatase protein and identifying and evaluating modulators of phosphatase protein activity. Other examples of transgenic animals

include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the phosphatase protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the phosphatase protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT

International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect substrate binding, kinase protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* phosphatase protein function, including substrate interaction, the effect of specific mutant phosphatase proteins on phosphatase protein function and substrate interaction, and the effect of chimeric phosphatase proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more phosphatase protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7;

(b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6; and

(d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7;

(b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6; and

(d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence

selected from the group consisting of SEQ ID NOS:2, 5, and 7, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

9. A host cell containing the vector of claim 8.

10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method

comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human phosphatase protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human phosphatase peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 5, and 7.

22. An isolated nucleic acid molecule encoding a human phosphatase peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6.

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, 4, and 6.

SPLICE FORM 1:

```
   1  CGTCCTTCCT GGTCCTGCGG GTCCAGGACT GTCCGCGGGG TTGAGGGAAG
  51  GGGCCGTGCC CGGTGCCAGC CCAGGTGCTC GCGGCCTGGC TCCATGGCCC
 101  TGGTCACAGT GAGCCGTTCG CCCCCGGGCA GCGGCGCCTC CACGCCCGTG
 151  GGGCCCTGGG ACCAGGCGGT CCAGCGAAGG AGTCGACTCC AGCGAAGGCA
 201  GAGCTTTGCG GTGCTCCGTG GGGCTGTCCT GGGACTGCAG GATGGAGGGG
 251  ACAATGATGA TGCAGCAGAG GCCAGTTCTG AGCCAACAGA GAAGGCCCCG
 301  AGTGAGGAGG AGCTCCACGG GGACCAGACA GACTTCGGGC AAGGATCCCA
 351  GAGTCCCCAG AAGCAGGAGG AGCAGAGGCA GCACCTGCAC CTCATGGTAC
 401  AGCTGCTGAG GCCGCAGGAT GACATCCGCC TGGCAGCCCA GCTGGAGGCA
 451  CCCCGGCCTC CCCGGCTCCG CTACCTGCTG GTAGTTTCTA CACGAGAAGG
 501  AGAAGGTCTG AGCCAGGATG AGACGGTCCT CCTGGGCGTG GATTTCCCTG
 551  ACAGCAGCTC CCCCAGCTGC ACCCTGGGCC TGGTCTTGCC CCTCTGGAGT
 601  GACACCCAGG TGTACTTAGA TGGAGACGGG GGCTTCAGCG TGACGTCTGG
 651  TGGGCAAAGC CGGATCTTCA AGCCCATCTC CATCCAGACC ATGTGGGCCA
 701  CACTCCAGGT ATTGCACCAA GCATGTGAGG CAGCTCTAGG CAGCGGCCTT
 751  GTACCGGGTG GCAGTGCCCT CACCTGGGCC AGCCACTACC AGGAGAGACT
 801  GAACTCCGAA CAGAGCTGCC TCAATGAGTG GACGGCTATG GCCGACCTGG
 851  AGTCTCTGCG GCCTCCCAGC GCCGAGCCTG GCGGGTCCTC AGAACAGGAG
 901  CAGATGGAGC AGGCGATCCG TGCTGAGCTG TGGAAAGTGT TGGATGTCAG
 951  TGACCTGGAG AGTGTCACTT CCAAAGAGAT CCGCCAGGCT CTGGAGCTGC
1001  GCCTGGGGCT CCCCCTCCAG CAGTACCGTG ACTTCATCGA CAACCAGATG
1051  CTGCTGCTGG TGGCACAGCG GGACCGAGCC TCCCGCATCT TCCCCCACCT
1101  CTACCTGGGC TCAGAGTGGA ACGCAGCAAA CCTGGAGGAG CTGCAGAGGA
1151  ACAGGGTCAC CCACATCTTG AACATGGCCC GGGAGATTGA CAACTTCTAC
1201  CCTGAGCGCT TCACCTACCA CAATGTGCGC CTCTGGGATG AGGAGTCGGC
1251  CCAGCTGCTG CCGCACTGGA AGGAGACGCA CCGCTTCATT GAGGCTGCAA
1301  GAGCACAGGG CACCCACGTG CTGGTCCACT GCAAGATGGG CGTCAGCCGC
1351  TCAGCGGCCA CAGTGCTGGC CTATGCCATG AAGCAGTACG AATGCAGCCT
1401  GGAGCAGGCC CTGCGCCACG TGCAGGAGCT CCGGCCCATC GCCCGCCCCA
1451  ACCCTGGCTT CCTGCGCCAG CTGCAGATCT ACCAGGGCAT CCTGACGGCC
1501  AGAACCTGAG GGTGGTGGGG AGGAGAAGGT TGTAGGCATG GAAGAGAGCC
1551  AGGCAGCCCC GAAAGAAGAG CCTGGGCCAC GGCCACGTAT AAACCTCCGA
1601  GGGGTCATGA GGTCCATCAG TCTTCTGGAG CCCTCCTTGG AGCTGGAGAG
1651  CACCTCAGAG ACCAGTGACA TGCCAGAGGT CTTCTCTTCC CACGAGTCTT
1701  CACATGAAGA GCCTCTGCAG CCCTTCCCAC AGCTTGCAAG GACCAAGGGA
1751  GGCCAGCAGG TGGACAGGGG GCCTCAGCCT GCCCTGAAGT CCCGCCAGTC
1801  AGTGGTTACC CTCCAGGGCA GTGCCGTGGT GGCCAACCGG ACCCAGGCCT
1851  TCCAGGAGCA GGAGCAGGGG CAGGGGCAGG GGCAGGGAGA GCCCTGCATT
1901  TCCTCTACGC CCAGGTTCCG GAAGGTGGTG AGACAGGCCA GCGTGCATGA
1951  CAGTGGAGAG GAGGGCGAGG CCTGAGCCCT CACACATGCC CACGCTCCCC
2001  TGACACTGAA GAGGATCCAC AACTCCTTGG AGAAACACCC TCACGTCTGT
2051  TGCCGCACAC ATTCCTCTCA GCTCCGCCCC ATACCCGTCA CTACAGCCTC
2101  ACCTCCCACC CCTGTCACTA CGGCCTCACC TCCCACCCCT GTCACTACAG
2151  CCTCACCTCC TACAGCCTTA AGTCCCAGGC CCATGTCTGC CTGTCCAAGG
2201  GCTCAAGACT TTCTAACTGG GATGTGGTAG AGGGACTGAA GGTACCTTTG
2251  GGGGCAACAG CACCCTAGTT TCATTCTCAA CTCTAGCCCT GCACACTCAC
2301  CTGTGGCACG GAATGAAAAC AGAGCTTCCC GTGCAAAAAG GGTCACGCCT
2351  CCCACCCCCG CCCCCTCCCT GCACCTCCTG TCCTCTCCCA GTTCATTCCT
2401  GGAACCAGCC AGGCCAGGCA ACCAGTGGCC CCCAAAGGCA GGCAGGATCC
2451  TCAGGCCCCA GCCGCGGGAG GCTGGAAGGG CTGGCAGATC GCTTCCCTCA
2501  TCCACCTCCA CCGGTCCAGG TCTTTGCTGC TGTCCCCAGA CCTCCTGTGA
2551  CACCACGCCA GATCACAGGG CACCAGGCCA GAGATAGTCT TCTTTTTGTC
2601  CTTTCTGGCC TCTGGCTAGT CAGTTTTTCA TAGCCTTACA GTATCTGGCT
2651  TTGTACTGAG AAATAAAACA CATTTTCATA AAAAAAAAAA AAAAAAAAAA
2701  AAAA (SEQ ID NO:1)
```

FEATURES:
5′ UTR: 1-93
Start: 94-1506
Stop: 1509
3′ UTR: 1510-2704

SPLICE FORM 2:

FIGURE 1A

```
   1  TGGTTGAGGG AAGGGGCCGT GCCCGGTGCC AGCCCAGGTG CTCGCGGCCT
  51  GGCTCCATGG CCCTGGTCAC AGTGAGCCGT TCGCCCCCGG GCAGCGGCGC
 101  CTCCACGCCC GTGGGGCCCT GGGACCAGGC GGTCCAGCGA AGGAGTCGAC
 151  TCCAGCGAAG GCAGAGCTTT GCGGTGCTCC GTGGGGCTGT CCTGGGACTG
 201  CAGGATGGAG GGGACAATGA TGATGCAGCA GAGGCCAGTT CTGAGCCAAC
 251  AGAGAAGGCC CCGAGTGAGG AGGAGCTCCA CGGGGACCAG ACAGACTTCG
 301  GGCAAGGATC CCAGAGTCCC CAGAAGCAGG AGGAGCAGAG GCAGCACCTG
 351  CACCTCATGG TACAGCTGCT GAGGCCGCAG GATGACATCC GCCTGGCAGC
 401  CCAGCTGGAG GCACCCCGGC CTCCCCGGCT CCGCTACCTG CTGGTAGTTT
 451  CTACACGAGA AGGAGAAGGT CTGAGCCAGG ATGAGACGGT CCTCCTGGGC
 501  GTGGATTTCC CTGACAGCAG CTCCCCCAGC TGCACCCTGG GCCTGGTCTT
 551  GCCCCTCTGG AGTGACACCC AGGTGTACTT AGATGGAGAC GGGGGCTTCA
 601  GCGTGACGTC TGGTGGGCAA AGCCGGATCT TCAAGCCCAT CTCCATCCAG
 651  ACCATGTGGG CCACACTCCA GGTATTGCAC CAAGCATGTG AGGCAGCTCT
 701  AGGCAGCGGC CTTGTACCGG GTGGCAGTGC CCTCACCTGG GCCAGCCACT
 751  ACCAGGAGAG ACTGAACTCC GAACAGAGCT GCCTCAATGA GTGGACGGCT
 801  ATGGCCGACC TGGAGTCTCT GCGGCCTCCC AGCGCCGAGC CTGGCGGGTC
 851  CTCAGAACAG GAGCAGATGG AGCAGGCGAT CCGTGCTGAG CTGTGGAAAG
 901  TGTTGGATGT CAGTGACCTG GAGAGTGTCA CTTCCAAAGA GATCCGCCAG
 951  GCTCTGGAGC TGCGCCTGGG GCTCCCCCTC CAGCAGTACC GTGACTTCAT
1001  CGACAACCAG ATGCTGCTGC TGGTGGCACA GCGGGACCGA GCCTCCCGCA
1051  TCTTCCCCCA CCTCTACCTG GGCTCAGAGT GGAACGCAGC AAACCTGGAG
1101  GAGCTGCAGA GGAACAGGGT CACCCACATC TTGAACATGG CCCGGGAGAT
1151  TGACAACTTC TACCCTGAGC GCTTCACCTA CCACAATGTG CGCCTCTGGG
1201  ATGAGGAGTC GGCCCAGCTG CTGCCGCACT GGAAGGAGAC GCACCGCTTC
1251  ATTGAGGCTG CAAGAGCACA GGGCACCCAC GTGCTGGTCC ACTGCAAGAT
1301  GGGCGTCAGC CGCTCAGCGG CCACAGTGCT GGCCTATGCC ATGAAGCAGT
1351  ACGAATGCAG CCTGGAGCAG GCCCTGCGCC ACGTGCAGGA GCTCCGGCCC
1401  ATCGCCCGCC CCAACCCTGG CTTCCTGCGC CAGCTGCAGA TCTACCAGGG
1451  CATCCTGACG GCCAGCCGCC AGAGCCATGT CTGGGAGCAG AAAGTGGGTG
1501  GGGTCTCCCC AGAGGAGCAC CCAGCCCCTG AAGTCTCTAC ACCATTCCCA
1551  CTTCTTCCGC CAGAACCTGA GGGTGGTGGG GAGGAGAAGG TTGTAGGCAT
1601  GGAAGAGAGC CAGGCAGCCC CGAAAGAAGA GCCTGGGCCA CGGCCACGTA
1651  TAAACCTCCG AGGGGTCATG AGGTCCATCA GTCTTCTGGA GCCCTCCTTG
1701  GAGCTGGAGA GCACCTCAGA GACCAGTGAC ATGCCAGAGG TCTTCTCTTC
1751  CCACGAGTCT TCACATGAAG AGCCTCTGCA GCCCTTCCCA CAGCTTGCAA
1801  GGACCAAGGG AGGCCAGCAG GTGGACAGGG GGCCTCAGCC TGCCCTGAAG
1851  TCCCGCCAGT CAGTGGTTAC CCTCCAGGGC AGTGCCGTGG TGGCCAACCG
1901  GACCCAGGCC TTCCAGGAGC AGGAGCAGGG GCAGGGGCAG GGGCAGGGAG
1951  AGCCCTGCAT TTCCTCTACG CCCAGGTTCC GGAAGGTGGT GAGACAGGCC
2001  AGCGTGCATG ACAGTGGAGA GGAGGGCGAG GCCTGAGCCC TCACACATGC
2051  CCACGCTCCC CTGACACTGA AGAGGATCCA CAACTCCTTG GAGAAACACC
2101  CTCACGTCTG TTGCCGCACA CATTCCTCTC AGCTCCGCCC CATACCCGTC
2151  ACTACAGCCT CACCTCCCAC CCCTGTCACT ACGGCCTCAC CTCCCACCCC
2201  TGTCACTACA GCCTCACCTC CTACAGCCTT AAGTCCCAGG CCCATGTCTG
2251  CCTGTCCAAG GGCTCAAGAC TTTCTAACTG GGATGTGGTA GAGGGACTGA
2301  AGGTACCTTT GGGGGCAACA GCACCCTAGT TTCATTCTCA ACTCTAGCCC
2351  TGCACACTCA CCTGTGGCAC GGAATGAAAA CAGAGCTTCC CGTGCAAAAA
2401  GGGTCACGCC TCCCACCCCC GCCCCCTCCC TGCACCTCCT GTCCTCTCCC
2451  AGTTCATTCC TGGAACCAGC CAGGCCAGGC AACCAGTGGC CCCCAAAGGC
2501  AGGCAGGATC CTCAGGCCCC AGCCGCGGGA GGCTGGAAGG GCTGGCAGAT
2551  CGCTTCCCTC ATCCACCTCC ACCGGTCCAG GTCTTTGCTG CTGTCCCCAG
2601  ACCTCCTGTG ACACCACGCC AGATCACAGG GCACCAGGCC AGAGATAGTC
2651  TTCTTTTTGT CCTTTCTGGC CTCTGGCTAG TCAGTTTTTC ATAGCCTTAC
2701  AGTATCTGGC TTTGTACTGA GAAATAAAAC ACATTTTCAT AAAAAAAAAA
2751  AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2801  AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
2851  AA (SEQ ID NO:2)
```

FEATURES:
5′ UTR: 1-56
Start: 57
Stop: 2034
3′ UTR: 2037-2852

FIGURE 1B

SPLICE FORM 3:

```
   1  CCTGGTCCTG CGGGTCCAGG ACTGTCCCGC GGGGTTGAGG GAAGGGGCCG
  51  TGCCCGGTGC CAGCCCAGGT GCTCGCGGCC TGGCTCCATG GCCCTGGTCA
 101  CAGTGAGCCG TTCGCCCCCG GGCAGCGGCG CCTCCACGCC CGTGGGGCCC
 151  TGGGACCAGG CGGTCCAGCG AAGGAGTCGA CTCCAGCGAA GGCAGAGCTT
 201  TGCGGTGCTC CGTGGGGCTG TCCTGGGACT GCAGGATGGA GGGGACAATG
 251  ATGATGCAGC AGAGGCCAGT TCTGAGCCAA CAGAGAAGGC CCCGAGTGAG
 301  GAGGAGCTCC ACGGGGACCA GACAGACTTC GGGCAAGGAT CCCAGAGTCC
 351  CCAGAAGCAG GAGGAGCAGA GGCAGCACCT GCACCTCATG GTACAGCTGC
 401  TGAGGCCGCA GGATGACATC CGCCTGGCAG CCCAGCTGGA GGCACCCCGG
 451  CCTCCCCGGC TCCGCTACCT GCTGGTAGTT TCTACACGAG AAGGAGAAGG
 501  TCTGAGCCAG GATGAGACGG TCCTCCTGGG CGTGGATTTC CCTGACAGCA
 551  GCTCCCCCAG CTGCACCCTG GGCCTGGTCT TGCCCCTCTG GAGTGACACC
 601  CAGGTGTACT TAGATGGAGA CGGGGGCTTC AGCGTGACGT CTGGTGGGCA
 651  AAGCCGGATC TTCAAGCCCA TCTCCATCCA GACCATGTGG TCCTCAGAAC
 701  AGGAGCAGAT GGAGCAGGCG ATCCGTGCTG AGCTGTGGAA AGTGTTGGAT
 751  GTCAGTGACC TGGAGAGTGT CACTTCCAAA GAGATCCGCC AGGCTCTGGA
 801  GCTGCGCCTG GGGCTCCCCC TCCAGCAGTA CCGTGACTTC ATCGACAACC
 851  AGATGCTGCT GCTGGTGGCA CAGCGGGACC GAGCCTCCCG CATCTTCCCC
 901  CACCTCTACC TGGGCTCAGA GTGGAACGCA GCAAACCTGG AGGAGCTGCA
 951  GAGGAACAGG GTCACCCACA TCTTGAACAT GGCCCGGGAG ATTGACAACT
1001  TCTACCCTGA GCGCTTCACC TACCACAATG TGCGCCTCTG GGATGAGGAG
1051  TCGGCCCAGC TGCTGCCGCA CTGGAAGGAG ACGCACCGCT TCATTGAGGC
1101  TGCAAGAGCA CAGGGCACCC ACGTGCTGGT CCACTGCAAG ATGGGCGTCA
1151  GCCGCTCAGC GGCCACAGTG CTGGCCTATG CCATGAAGCA GTACGAATGC
1201  AGCCTGGAGC AGGCCCTGCG CCACGTGCAG GAGCTCCGGC CCATCGCCCG
1251  CCCCAACCCT GGCTTCCTGC GCCAGCTGCA GATCTACCAG GGCATCCTGA
1301  CGGCCAGAAC CTGAGGGTGG TGGGGAGGAG AAGGTTGTAG GCATGGAAGA
1351  GAGCCAGGCA GCCCCGAAAG AAGAGCCTGG GGCCACGGGG CACGTATAAA
1401  CCTCCGAGGG GTCATGAGGT CCATCAGTCT TCTGGAGCCC TCCTTGGGAG
1451  CTGGAGAGCA CCTCAGTAGA CCAGTGACAT GCCAGAGGTC TTCTCTTCCC
1501  ACGAGTCTTC ACATGAAGAG CCTCTGCAGC CCTTCCCACA GCTTGCAAGG
1551  ACCAAGGGAG GCCAGCAGGT GGACAGGGGG CCTCAGCCTG CCCTGAAGTC
1601  CCGCCAGTCA GTGGTTACCC TCCAGGGCAG TGCCGTGGTG GCCAACCGGA
1651  CCCAGGCCTT CCAGGAGCAG GAGCAGGGGC AGGGGCAGGG GCAGGGAGAG
1701  CCCTGCATTT CCTCTACGCC CAGGTTCCGG AAGGTGGTGA GACAGGCCAG
1751  CGTGCATGAC AGTGGAGAGG AGGGCGAGGC CTGAGCCCTC ACACATGCCC
1801  ACGCTCCCCT GACACTGAAG AGGATCCACA ACTCCTTGGA GAAACACCCT
1851  CACGTCTGTT GCCGCACACA TTCCTCTCAG CTCCGCCCCA TACCCGTCAC
1901  TACAGCCTCA CCTCCCACCC CTGTCACTAC GGCCTCACCT CCCACCCCTG
1951  TCACTACAGC CTCACCTCCT ACAGCCTTAA GTCCCAGGCC CATGTCTGCC
2001  TGTCCAAGGG CTCAAGACTT TCTAACTGGG ATGTGGTAGA GGGACTGAAG
2051  GTACCTTTGG GGGCAACAGC ACCCTAGTTT CATTCTCAAC TCTAGCCCTG
2101  CACACTCACC TGTGGCACGG AATGAAAACA GAGCTTCCCG TGCAAAAAGG
2151  GTCACGCCTC CCACCCCCGC CCCCTCCCTG CACCTCCTGT CCTCTCCCAG
2201  TTCATTCCTG GAACCAGCCA GGCCAGGCAA CCAGTGGCCC CCAAAGGCAG
2251  GCAGGATCCT CAGGCCCCAG CCGCGGGAGG CTGGAAGGGC TGGCAGATCG
2301  CTTCCCTCAT CCACCTCCAC CGGTCCAGGT CTTTGCTGCT GTCCCCAGAC
2351  CTCCTGTGAC ACCACGCCAG ATCACAGGGC ACCAGGCCAG AGATAGTCTT
2401  CTTTTTGTCC TTTCTGGCCT CTGGCTAGTC AGTTTTTCAT AGCCTTACAG
2451  TATCTGGCTT TGTACTGAGA AATAAAACAC ATTTTCATAT TTGGTTAAAA
2501  AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA (SEQ ID NO:3)
```

FEATURES:
5′ UTR: 1-88
Start: 88
Stop: 1311
3′ UTR: 1315-2540

Homologous proteins:
Top 10 BLAST Hits

FIGURE 1C

	Score	E
SPLICE FORM 1:		
gi|8923483|ref|NP_060327.1| hypothetical protein FLJ20515 >gi|7...	576	e-163
gi|6714641|dbj|BAA89534.1| (AB036834) MAP kinase phosphatase [D...	337	2e-91
gi|8922777|ref|NP_060746.1| hypothetical protein FLJ10928 >gi|7...	233	3e-60
gi|7242951|dbj|BAA92536.1| (AB037719) KIAA1298 protein [Homo sa...	143	4e-33
gi|7301242|gb|AAF56372.1| (AE003750) CG6238 gene product [Droso...	124	2e-27
gi|4150963|emb|CAA77232.1| (Y18620) DsPTP1 protein [Arabidopsis...	113	5e-24
gi|9294518|dbj|BAB02780.1| (AB023036) dual-specificity protein ...	113	5e-24
gi|6862915|gb|AAF30304.1|AC018907_4 (AC018907) putative dual-sp...	94	2e-18
gi|6015037|sp|O54838|DUS5_RAT DUAL SPECIFICITY PROTEIN PHOSPHAT...	92	9e-18
gi|9910432|ref|NP_064570.1| mitogen-activated protein kinase ph...	90	3e-17
SPLICE FORM 2:		
gi|8923483|ref|NP_060327.1| hypothetical protein FLJ20515 [Homo...	576	e-163
gi|6714641|dbj|BAA89534.1| (AB036834) MAP kinase phosphatase [D...	340	6e-92
gi|8922777|ref|NP_060746.1| hypothetical protein FLJ10928 [Homo...	229	1e-58
gi|7242951|dbj|BAA92536.1| (AB037719) KIAA1298 protein [Homo sa...	162	1e-38
gi|9294518|dbj|BAB02780.1| (AB023036) dual-specificity protein ...	113	8e-24
gi|4150963|emb|CAA77232.1| (Y18620) DsPTP1 protein [Arabidopsis...	113	8e-24
gi|4758212|ref|NP_004411.1| dual specificity phosphatase 8 [Hom...	99	3e-19
gi|6679156|ref|NP_032774.1| neuronal tyrosine/threonine phospha...	96	2e-18
gi|6862915|gb|AAF30304.1|AC018907_4 (AC018907) putative dual-sp...	94	6e-18
gi|6015037|sp|O54838|DUS5_RAT DUAL SPECIFICITY PROTEIN PHOSPHAT...	92	2e-17
SPLICE FORM 3:		
gi|8923483|ref|NP_060327.1| hypothetical protein FLJ20515 [Homo...	410	e-113
gi|8922777|ref|NP_060746.1| hypothetical protein FLJ10928 [Homo...	233	7e-60
gi|6714641|dbj|BAA89534.1| (AB036834) MAP kinase phosphatase [D...	224	5e-57
gi|7242951|dbj|BAA92536.1| (AB037719) KIAA1298 protein [Homo sa...	143	5e-33
gi|9294518|dbj|BAB02780.1| (AB023036) dual-specificity protein ...	113	1e-23
gi|4150963|emb|CAA77232.1| (Y18620) DsPTP1 protein [Arabidopsis...	113	1e-23
gi|6862915|gb|AAF30304.1|AC018907_4 (AC018907) putative dual-sp...	94	5e-18
gi|6015037|sp|O54838|DUS5_RAT DUAL SPECIFICITY PROTEIN PHOSPHAT...	92	2e-17
gi|9910432|ref|NP_064570.1| mitogen-activated protein kinase ph...	90	7e-17
gi|9911130|gb|AAA64693.2| (U15932) protein phosphatase [Homo sa...	90	1e-16

BLAST to dbEST:

SPLICE FORM 1:

	Score	E
gi|9807071 /dataset=dbest /taxon=960...	1404	0.0
gi|10317998 /dataset=dbest /taxon=96...	1316	0.0
gi|10151079 /dataset=dbest /taxon=96...	1249	0.0
gi|10401153 /dataset=dbest /taxon=960...	1180	0.0
gi|10329921 /dataset=dbest /taxon=96...	1124	0.0
gi|7632969 /dataset=dbest /taxon=960...	791	0.0
gi|9155111 /dataset=dbest /taxon=9606...	779	0.0
gi|10994242 /dataset=dbest /taxon=96...	450	e-124

EXPRESSION INFORMATION FOR MODULATORY USE:

SPLICE FORM 1:

library source:

Expression information from BLAST dbEST hits:

gi|9807071	Human Pancreas
gi|10317998	Human colon adenocarcinoma
gi|10151079	Human Pancreas:adenocarcinoma
gi|10401153	Human Pancreas:epitheliod carcinoma
gi|10329921	Human lung: large cell carcinoma
gi|7632969	Human kidney: renal cell carcinoma
gi|9155111	Human Placenta choriocarcinoma
gi|10994242	Human ovary tumor tissue

FIGURE 1D

Expression information from PCR-based tissue screening panels:
Human Brain
Human Fetal brain
Human fetal heart
Human fetal kidney
Human heart
Human kidney
Human uterus
Human thyroid

FIGURE 1E

SPLICE FORM 1:

```
   1 MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD
  51 GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL
 101 MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD
 151 FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM
 201 WATLQVLHQA CEAALGSGLV PGGSALTWAS HYQERLNSEQ SCLNEWTAMA
 251 DLESLRPPSA EPGGSSEQEQ MEQAIRAELW KVLDVSDLES VTSKEIRQAL
 301 ELRLGLPLQQ YRDFIDNQML LLVAQRDRAS RIFPHLYLGS EWNAANLEEL
 351 QRNRVTHILN MAREIDNFYP ERFTYHNVRL WDEESAQLLP HWKETHRFIE
 401 AARAQGTHVL VHCKMGVSRS AATVLAYAMK QYECSLEQAL RHVQELRPIA
 451 RPNPGFLRQL QIYQGILTAR T (SEQ ID NO:4)
```

SPLICE FORM 2:

```
   1  MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD
  51  GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL
 101  MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD
 151  FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM
 201  WATLQVLHQA CEAALGSGLV PGGSALTWAS HYQERLNSEQ SCLNEWTAMA
 251  DLESLRPPSA EPGGSSEQEQ MEQAIRAELW KVLDVSDLES VTSKEIRQAL
 301  ELRLGLPLQQ YRDFIDNQML LLVAQRDRAS RIFPHLYLGS EWNAANLEEL
 351  QRNRVTHILN MAREIDNFYP ERFTYHNVRL WDEESAQLLP HWKETHRFIE
 401  AARAQGTHVL VHCKMGVSRS AATVLAYAMK QYECSLEQAL RHVQELRPIA
 451  RPNPGFLRQL QIYQGILTAS RQSHVWEQKV GGVSPEEHPA PEVSTPFPLL
 501  PPEPEGGGEE KVVGMEESQA APKEEPGPRP RINLRGVMRS ISLLEPSLEL
 551  ESTSETSDMP EVFSSHESSH EEPLQPFPQL ARTKGGQQVD RGPQPALKSR
 601  QSVVTLQGSA VVANRTQAFQ EQEQGQGQGQ GEPCISSTPR FRKVVRQASV
 651  HDSGEEGEA (SEQ ID NO:5)
```

SPLICE FORM 3:

```
   1  MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD
  51  GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL
 101  MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD
 151  FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM
 201  WSSEQEQMEQ AIRAELWKVL DVSDLESVTS KEIRQALELR LGLPLQQYRD
 251  FIDNQMLLLV AQRDRASRIF PHLYLGSEWN AANLEELQRN RVTHILNMAR
 301  EIDNFYPERF TYHNVRLWDE ESAQLLPHWK ETHRFIEAAR AQGTHVLVHC
 351  KMGVSRSAAT VLAYAMKQYE CSLEQALRHV QELRPIARPN PGFLRQLQIY
 401  QGILTART  (SEQ ID NO:6)
```

FEATURES:
Functional domains and key regions:
SPLICE FORM 1:
[1] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

34-37 RRQS

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 6

1	65-67 TEK
2	132-134 STR
3	254-256 SLR
4	292-294 TSK
5	395-397 THR
6	468-470 TAR

FIGURE 2A

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 6

1	70-73	SEEE
2	132-135	STRE
3	140-143	SQDE
4	266-269	SEQE
5	286-289	SDLE
6	292-295	TSKE

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

363-369 REIDNFY

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 8

1	12-17	GSGAST
2	43-48	GAVLGL
3	47-52	GLQDGG
4	218-223	GLVPGG
5	223-228	GSALTW
6	339-344	GSEWNA
7	416-421	GVSRSA
8	465-470	GILTAR

FIGURE 2B

BLAST Alignment to Top Hit:
SPLICE FORM 1:

```
>gi|8923483|ref|NP_060327.1| hypothetical protein FLJ20515
          >gi|7020674|dbj|BAA91228.1| (AK000522) unnamed protein
          product [Homo sapiens]
          Length = 394

 Score =  576 bits (1469), Expect = e-163
 Identities = 290/312 (92%), Positives = 296/312 (93%), Gaps = 1/312 (0%)

Query: 1   MALVTVSRSPPGSGASTPVGPWDQAVQRRSRLQRRQSFAVLRGAVLGLQDGGDNDDAAEA 60
           MALVTVSRSPPGSGASTPVGPWDQAVQRRSRLQRRQSFAVLRGAVLGLQDGGDNDDAAEA
Sbjct: 1   MALVTVSRSPPGSGASTPVGPWDQAVQRRSRLQRRQSFAVLRGAVLGLQDGGDNDDAAEA 60

Query: 61  SSEPTEKAPSEEELHGDQTDFGQGSQSPQKQEEQRQHLHLMVQLLRPQDDIRLAAQLEAP 120
           SSEPTEKAPSEEELHGDQTDFGQGSQSPQKQEEQRQHLHLMVQLLRPQDDIRLAAQLEAP
Sbjct: 61  SSEPTEKAPSEEELHGDQTDFGQGSQSPQKQEEQRQHLHLMVQLLRPQDDIRLAAQLEAP 120

Query: 121 RPPRLRYLLVVSTREGEGLSQDETVLLGVDFPDSSSPSCTLGLVLPLWSDTQVYLDGDGG 180
           RPPRLRYLLVVSTREGEGLSQDETVLLGVDFPDSSSPSCTLGLVLPLWSDTQVYLDGDGG
Sbjct: 121 RPPRLRYLLVVSTREGEGLSQDETVLLGVDFPDSSSPSCTLGLVLPLWSDTQVYLDGDGG 180

Query: 181 FSVTSGGQSRIFKPISIQTMWATLQVLHQACEAALGSGLVPGGSALTWASHYQERLNSEQ 240
           FSVTSGGQSRIFKPISIQTMWATLQVLHQACEAALGSGLVPGGSALTWASHYQERLNSEQ
Sbjct: 181 FSVTSGGQSRIFKPISIQTMWATLQVLHQACEAALGSGLVPGGSALTWASHYQERLNSEQ 240

Query: 241 SCLNEWTAMADLESLRPPSAEPGGSSEQEQMEQAIRAELWKVLDV-SDLESVTSKEIRQA 299
           SCLNEWTAMADLESLRPPSAEPGGSSEQEQMEQAIRAELWKVL++ S  E+    E+  +
Sbjct: 241 SCLNEWTAMADLESLRPPSAEPGGSSEQEQMEQAIRAELWKVLELESTSETSDMPEVFSS 300

Query: 300 LELRLGLPLQQY 311
            E     PLQ +
Sbjct: 301 HESSHEEPLQPF 312 (SEQ ID NO:8)
```

```
>gi|6714641|dbj|BAA89534.1| (AB036834) MAP kinase phosphatase
          [Drosophila melanogaster]
          Length = 1045

 Score =  337 bits (854), Expect = 2e-91
 Identities = 204/537 (37%), Positives = 284/537 (51%), Gaps = 81/537 (15%)

Query: 1   MALVTVSRSPPGSGA-STPVGPWDQAVQRRSRLQRRQSFAVLRGAVLGLQDGGDNDDAAE 59
           MALVTV RSP  +G+ S   G  +          R + F   +G  L L
Sbjct: 1   MALVTVQRSPSVAGSCSNSDGESEDDEGNSKGNDRSECFFAGKGTALVL---------- 49

Query: 60  ASSEPTEKAPSEEELHGDQTDFGQGSQSPQKQEEQRQHLHLMVQLLRPQDDIRLAAQLEA 119
           A  +      SE  L  D T     +QS     + + HL  M  LL+ +D +++A +LE+
Sbjct: 50  ALKDIPPLTQSERRLSTDSTRSSNSTQS--NNSDIQLHLQSMFYLLQREDTLKMAVKLES 107

Query: 120 PRPPRLRYLLV---------VSTREGEGLSQDETVLLGVDFPDSSSPS---------- 158
            R  R RYL++          S R    + +  +V +G      SS S
Sbjct: 108 QRSNRTRYLVIASRSCCRSGTSDRRRHRIMRHHSVKVGGSAGTKSSTSPAVPTQRQLSVE 167

Query: 159 ---------------------------------CTLGL----------VLPLWSDTQVY 174
                                            C LG+           V+P+ +DT ++
Sbjct: 168 QTATEASSKCDKTADKENATAAGDNKNTSGMEESCLLGIDCNERTTIGLVVPILADTTIH 227

Query: 175 LDGDGGFSVTSGGQSRIFKPISIQTMWATLQVLHQACEAALGSGLVPGGSALTWASHYQE 234
           LDGDGGFSV    ++ IFKP+S+Q MW+ LQ LH+  + A  +     G +  W S Y+
Sbjct: 228 LDGDGGFSVKVYEKTHIFKPVSVQAMWSALQTLHKVSKKARENNFYASGPSHDWLSSYER 287
```

FIGURE 2C

```
Query: 235 RLNSEQSCLNEWTAMADLESLRPPSAEP--GGSSEQEQMEQAIRAELWKVLDVSDLESVT 292
           R+ S+QSCLNEW AM  LES RPPS +       E+E+ E  I+ +L  ++   DL+ VT
Sbjct: 288 RIESDQSCLNEWNAMDALESRRPPSPDAIRNKPPEKEETESVIKMKLKAIMMSVDLDEVT 347

Query: 293 SKEIRQALELRLGLPLQQYRDFIDNQMLLLVAQRDRASRIFPHLYLGSEWNAANLEELQR 352
           SK IR  LE  L + L +Y+ FID +ML+++ Q D  ++IF H+YLGSEWNA+NLEELQ+
Sbjct: 348 SKYIRGRLEEILDMDLGEYKSFIDAEMLVILGQMDAPTKIFEHVYLGSEWNASNLEELQK 407

Query: 353 NRVTHILNMAREIDNFYPERFTYHNVRLWDEESAQLLPHWKETHRFIEAARAQGTHVLVH 412
           N V HILN+ REIDNF+P  F Y NVR++D+E   LL +W +T R+I  A+A+G+ VLVH
Sbjct: 408 NGVRHILNVTREIDNFFPGTFEYFNVRVYDDEKTNLLKYWDDTFRYITRAKAEGSKVLVH 467

Query: 413 CKMGVSRSAATVLAYAMKQYECSLEQALRHVQELRPIARPNPGFLRQLQIYQGILTA 469
           CKMGVSRSA+ V+AYAMK Y+   +QAL HV++ R   +PN  FL QL+ Y G+L A
Sbjct: 468 CKMGVSRSASVVIAYAMKAYQWEFQQALEHVKKRRSCIKPNKNFLNQLETYSGMLDA 524
 (SEQ ID NO:9)

>gi|8922777|ref|NP_060746.1| hypothetical protein FLJ10928
          >gi|7023283|dbj|BAA91913.1| (AK001790) unnamed protein
          product [Homo sapiens]
          Length = 141

 Score =  233 bits (588), Expect = 3e-60
 Identities = 111/111 (100%), Positives = 111/111 (100%)

Query: 361 MAREIDNFYPERFTYHNVRLWDEESAQLLPHWKETHRFIEAARAQGTHVLVHCKMGVSRS 420
           MAREIDNFYPERFTYHNVRLWDEESAQLLPHWKETHRFIEAARAQGTHVLVHCKMGVSRS
Sbjct: 31  MAREIDNFYPERFTYHNVRLWDEESAQLLPHWKETHRFIEAARAQGTHVLVHCKMGVSRS 90

Query: 421 AATVLAYAMKQYECSLEQALRHVQELRPIARPNPGFLRQLQIYQGILTART 471
           AATVLAYAMKQYECSLEQALRHVQELRPIARPNPGFLRQLQIYQGILTART
Sbjct: 91  AATVLAYAMKQYECSLEQALRHVQELRPIARPNPGFLRQLQIYQGILTART 141
(SEQ ID NO:10)
```

FIGURE 2D

```
   1 CGTCCTTCCT GGTCCTGCGG GTCCAGGACT GTCCGCGGGG TTGAGGGAAG
  51 GGGCCGTGCC CGGTGCCAGC CCAGGTGCTC GCGGCCTGGC TCCATGGCCC
 101 TGGTCACAGT GAGCCGTTCG CCCCCGGGCA GCGGCGCCTC CACGCCCGTG
 151 GGGCCCTGGG ACCAGGCGGT CCAGCGAAGG AGTCGACTCC AGCGAAGGCA
 201 GAGCTTTGCG GTGCTCCGTG GGGCTGTCCT GGGACTGCAG GATGGAGGGG
 251 ACAATGATGA TGCAGCAGAG GCCAGTTCTG AGCCAACAGA GAAGGCCCCG
 301 AGTGAGGAGG AGCTCCACGG GGACCAGACA GACTTCGGGC AAGGATCCCA
 351 GAGTCCCCAG AAGCAGGAGG AGCAGAGGCA GCACCTGCAC CTCATGGTAC
 401 AGCTGCTGAG GCCGCAGGAT GACATCCGCC TGGCAGCCCA GCTGGAGGCA
 451 CCCCGGCCTC CCCGGCTCCG CTACCTGCTG GTAGTTTCTA CACGAGAAGG
 501 AGAAGGTCTG AGCCAGGATG AGACGGTCCT CCTGGGCGTG GATTTCCCTG
 551 ACAGCAGCTC CCCCAGCTGC ACCCTGGGCC TGGTCTTGCC CCTCTGGAGT
 601 GACACCCAGG TGTACTTAGA TGGAGACGGG GGCTTCAGCG TGACGTCTGG
 651 TGGGCAAAGC CGGATCTTCA AGCCCATCTC CATCCAGACC ATGTGGGCCA
 701 CACTCCAGGT ATTGCACCAA GCATGTGAGG CAGCTCTAGG CAGCGGCCTT
 751 GTACCGGGTG GCAGTGCCCT CACCTGGGCC AGCCACTACC AGGAGAGACT
 801 GAACTCCGAA CAGAGCTGCC TCAATGAGTG GACGGCTATG GCCGACCTGG
 851 AGTCTCTGCG GCCTCCCAGC GCCGAGCCTG GCGGGTCCTC AGAACAGGAG
 901 CAGATGGAGC AGGCGATCCG TGCTGAGCTG TGGAAAGTGT TGGATGTCAG
 951 TGACCTGGAG AGTGTCACTT CCAAAGAGAT CCGCCAGGCT CTGGAGCTGC
1001 GCCTGGGGCT CCCCCTCCAG CAGTACCGTG ACTTCATCGA CAACCAGATG
1051 CTGCTGCTGG TGGCACAGCG GGACCGAGCC TCCCGCATCT TCCCCCACCT
1101 CTACCTGGGC TCAGAGTGGA ACGCAGCAAA CCTGGAGGAG CTGCAGAGGA
1151 ACAGGGTCAC CCACATCTTG AACATGGCCC GGGAGATTGA CAACTTCTAC
1201 CCTGAGCGCT TCACCTACCA CAATGTGCGC CTCTGGGATG AGGAGTCGGC
1251 CCAGCTGCTG CCGCACTGGA AGGAGACGCA CCGCTTCATT GAGGCTGCAA
1301 GAGCACAGGG CACCCACGTG CTGGTCCACT GCAAGATGGG CGTCAGCCGC
1351 TCAGCGGCCA CAGTGCTGGC CTATGCCATG AAGCAGTACG AATGCAGCCT
1401 GGAGCAGGCC CTGCGCCACG TGCAGGAGCT CCGGCCCATC GCCCGCCCCA
1451 ACCCTGGCTT CCTGCGCCAG CTGCAGATCT ACCAGGGCAT CCTGACGGCC
1501 AGAACCTGAG GGTGGTGGGG AGGAGAAGGT TGTAGGCATG GAAGAGAGCC
1551 AGGCAGCCCC GAAAGAAGAG CCTGGGCCAC GGCCACGTAT AAACCTCCGA
1601 GGGGTCATGA GGTCCATCAG TCTTCTGGAG CCCTCCTTGG AGCTGGAGAG
1651 CACCTCAGAG ACCAGTGACA TGCCAGAGGT CTTCTCTTCC CACGAGTCTT
1701 CACATGAAGA GCCTCTGCAG CCCTTCCCAC AGCTTGCAAG GACCAAGGGA
1751 GGCCAGCAGG TGGACAGGGG GCCTCAGCCT GCCCTGAAGT CCCGCCAGTC
1801 AGTGGTTACC CTCCAGGGCA GTGCCGTGGT GGCCAACCGG ACCCAGGCCT
1851 TCCAGGAGCA GGAGCAGGGG CAGGGGCAGG GGCAGGGAGA GCCCTGCATT
1901 TCCTCTACGC CCAGGTTCCG GAAGGTGGTG AGACAGGCCA GCGTGCATGA
1951 CAGTGGAGAG GAGGGCGAGG CCTGAGCCCT CACACATGCC CACGCTCCCC
2001 TGACACTGAA GAGGATCCAC AACTCCTTGG AGAAACACCC TCACGTCTGT
2051 TGCCGCACAC ATTCCTCTCA GCTCCGCCCC ATACCCGTCA CTACAGCCTC
2101 ACCTCCCACC CCTGTCACTA CGGCCTCACC TCCCACCCCT GTCACTACAG
2151 CCTCACCTCC TACAGCCTTA AGTCCCAGGC CCATGTCTGC CTGTCCAAGG
2201 GCTCAAGACT TTCTAACTGG GATGTGGTAG AGGGACTGAA GGTACCTTTG
2251 GGGGCAACAG CACCCTAGTT TCATTCTCAA CTCTAGCCCT GCACACTCAC
2301 CTGTGGCACG GAATGAAAAC AGAGCTTCCC GTGCAAAAAG GGTCACGCCT
2351 CCCACCCCCG CCCCCTCCCT GCACCTCCTG TCCTCTCCCA GTTCATTCCT
2401 GGAACCAGCC AGGCCAGGCA ACCAGTGGCC CCCAAAGGCA GGCAGGATCC
2451 TCAGGCCCCA GCCGCGGGAG GCTGGAAGGG CTGGCAGATC GCTTCCCTCA
2501 TCCACCTCCA CCGGTCCAGG TCTTTGCTGC TGTCCCCAGA CCTCCTGTGA
2551 CACCACGCCA GATCACAGGG CACCAGGCCA GAGATAGTCT TCTTTTTGTC
2601 CTTTCTGGCC TCTGGCTAGT CAGTTTTTCA TAGCCTTACA GTATCTGGCT
2651 TTGTACTGAG AAATAAAACA CATTTTCATA AAAAAAAAAA AAAAAAAAAA
2701 AAAA (SEQ ID NO:7)
```

FEATURES:
Start: 94
Exon: 94-1506
Stop: 1507

CHROMOSOME MAP POSITION:
Bac accession number: AP001885
Chromosome #: 11

FIGURE 3A

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
577	G	A	Exon	162	G	S
1451	G	A	Exon	453	S	N
2641	G	A	Beyond ORF(3')			

Context:

DNA Position

577
```
TCTGAGCCAACAGAGAAGGCCCCGAGTGAGGAGGAGCTCCACGGGGACCAGACAGACTTC
GGGCAAGGATCCCAGAGTCCCCAGAAGCAGGAGGAGCAGAGGCAGCACCTGCACCTCATG
GTACAGCTGCTGAGGCCGCAGGATGACATCCGCCTGGCAGCCCAGCTGGAGGCACCCCGG
CCTCCCCGGCTCCGCTACCTGCTGGTAGTTTCTACACGAGAAGGAGAAGGTCTGAGCCAG
GATGAGACGGTCCTCCTGGGCGTGGATTTCCCTGACAGCAGCTCCCCCAGCTGCACCCTG
[G,A]
GCCTGGTCTTGCCCCTCTGGAGTGACACCCAGGTGTACTTAGATGGAGACGGGGGCTTCA
GCGTGACGTCTGGTGGGCAAAGCCGGATCTTCAAGCCCATCTCCATCCAGACCATGTGGG
CCACACTCCAGGTATTGCACCAAGCATGTGAGGCAGCTCTAGGCAGCGGCCTTGTACCGG
GTGGCAGTGCCCTCACCTGGGCCAGCCACTACCAGGAGAGACTGAACTCCGAACAGAGCT
GCCTCAATGAGTGGACGGCTATGGCCGACCTGGAGTCTCTGCGGCCTCCCAGCGCCGAGC
```

1451
```
ACAGGGTCACCCACATCTTGAACATGGCCCGGGAGATTGACAACTTCTACCCTGAGCGCT
TCACCTACCACAATGTGCGCCTCTGGGATGAGGAGTCGGCCCAGCTGCTGCCGCACTGGA
AGGAGACGCACCGCTTCATTGAGGCTGCAAGAGCACAGGGCACCCACGTGCTGGTCCACT
GCAAGATGGGCGTCAGCCGCTCAGCGGCCACAGTGCTGGCCTATGCCATGAAGCAGTACG
AATGCAGCCTGGAGCAGGCCCTGCGCCACGTGCAGGAGCTCCGGCCCATCGCCCGCCCCA
[G,A]
CCCTGGCTTCCTGCGCCAGCTGCAGATCTACCAGGGCATCCTGACGGCCAGAACCTGAGG
GTGGTGGGGAGGAGAAGGTTGTAGGCATGGAAGAGAGCCAGGCAGCCCCGAAAGAAGAGC
CTGGGCCACGGCCACGTATAAACCTCCGAGGGGTCATGAGGTCCATCAGTCTTCTGGAGC
CCTCCTTGGAGCTGGAGAGCACCTCAGAGACCAGTGACATGCCAGAGGTCTTCTCTTCCC
ACGAGTCTTCACATGAAGAGCCTCTGCAGCCCTTCCCACAGCTTGCAAGGACCAAGGGAG
```

2641
```
GGTCACGCCTCCCACCCCCGCCCCCTCCCTGCACCTCCTGTCCTCTCCCAGTTCATTCCT
GGAACCAGCCAGGCCAGGCAACCAGTGGCCCCCAAAGGCAGGCAGGATCCTCAGGCCCCA
GCCGCGGGAGGCTGGAAGGGCTGGCAGATCGCTTCCCTCATCCACCTCCACCGGTCCAGG
TCTTTGCTGCTGTCCCCAGACCTCCTGTGACACCACGCCAGATCACAGGGCACCAGGCCA
GAGATAGTCTTCTTTTTGTCCTTTCTGGCCTCTGGCTAGTCAGTTTTTCATAGCCTTACA
[G,A]
TATCTGGCTTTGTACTGAGAAATAAAACACATTTTCATAAAAAAAAAAAAAAAAAAAAAA
AAA
```

GENE STRUCTURE MODEL:

```
SPLICE FORM 2       5'-e1-e2-e3-e4-e5-e6-e7-e8-e9-e10-e11-e12-e13-e14-3'
SPLICE FORM 1       5'-e1-e2-e3-e4-e5-e6-e7-e8-e9-e10-e11-e12---------3'
SPLICE FORM 3       5'-e1-e2-e3-e4-e5-e6----e8-e9-e10-e11-e12---------3'
```

MULTIPLE ALIGNMENT OF CDNA SEQUENCES:

```
SPLICE FORM 2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~TG GTTGAGGGAA
SPLICE FORM 1 CGTCCTTCCT GGTCCTGCGG GTCCAGGACT GT.CCGCGGG GTTGAGGGAA
SPLICE FORM 3 ~~~~~~~~CCT GGTCCTGCGG GTCCAGGACT GTCCCGCGGG GTTGAGGGAA

              51                                                  100
SPLICE FORM 2 GGGGCCGTGC CCGGTGCCAG CCCAGGTGCT CGCGGCCTGG CTCCATGGCC
SPLICE FORM 1 GGGGCCGTGC CCGGTGCCAG CCCAGGTGCT CGCGGCCTGG CTCCATGGCC
SPLICE FORM 3 GGGGCCGTGC CCGGTGCCAG CCCAGGTGCT CGCGGCCTGG CTCCATGGCC

              101                                                 150
SPLICE FORM 2 CTGGTCACAG TGAGCCGTTC GCCCCCGGGC AGCGGCGCCT CCACGCCCGT
SPLICE FORM 1 CTGGTCACAG TGAGCCGTTC GCCCCCGGGC AGCGGCGCCT CCACGCCCGT
```

FIGURE 3B

```
SPLICE FORM 3 CTGGTCACAG TGAGCCGTTC GCCCCCGGGC AGCGGCGCCT CCACGCCCGT

              151                                                 200
SPLICE FORM 2 GGGGCCCTGG GACCAGGCGG TCCAGCGAAG GAGTCGACTC CAGCGAAGGC
SPLICE FORM 1 GGGGCCCTGG GACCAGGCGG TCCAGCGAAG GAGTCGACTC CAGCGAAGGC
SPLICE FORM 3 GGGGCCCTGG GACCAGGCGG TCCAGCGAAG GAGTCGACTC CAGCGAAGGC

              201                                                 250
SPLICE FORM 2 AGAGCTTTGC GGTGCTCCGT GGGGCTGTCC TGGGACTGCA GGATGGAGGG
SPLICE FORM 1 AGAGCTTTGC GGTGCTCCGT GGGGCTGTCC TGGGACTGCA GGATGGAGGG
SPLICE FORM 3 AGAGCTTTGC GGTGCTCCGT GGGGCTGTCC TGGGACTGCA GGATGGAGGG

              251                                                 300
SPLICE FORM 2 GACAATGATG ATGCAGCAGA GGCCAGTTCT GAGCCAACAG AGAAGGCCCC
SPLICE FORM 1 GACAATGATG ATGCAGCAGA GGCCAGTTCT GAGCCAACAG AGAAGGCCCC
SPLICE FORM 3 GACAATGATG ATGCAGCAGA GGCCAGTTCT GAGCCAACAG AGAAGGCCCC

              301                                                 350
SPLICE FORM 2 GAGTGAGGAG GAGCTCCACG GGGACCAGAC AGACTTCGGG CAAGGATCCC
SPLICE FORM 1 GAGTGAGGAG GAGCTCCACG GGGACCAGAC AGACTTCGGG CAAGGATCCC
SPLICE FORM 3 GAGTGAGGAG GAGCTCCACG GGGACCAGAC AGACTTCGGG CAAGGATCCC

              351                                                 400
SPLICE FORM 2 AGAGTCCCCA GAAGCAGGAG GAGCAGAGGC AGCACCTGCA CCTCATGGTA
SPLICE FORM 1 AGAGTCCCCA GAAGCAGGAG GAGCAGAGGC AGCACCTGCA CCTCATGGTA
SPLICE FORM 3 AGAGTCCCCA GAAGCAGGAG GAGCAGAGGC AGCACCTGCA CCTCATGGTA

              401                                                 450
SPLICE FORM 2 CAGCTGCTGA GGCCGCAGGA TGACATCCGC CTGGCAGCCC AGCTGGAGGC
SPLICE FORM 1 CAGCTGCTGA GGCCGCAGGA TGACATCCGC CTGGCAGCCC AGCTGGAGGC
SPLICE FORM 3 CAGCTGCTGA GGCCGCAGGA TGACATCCGC CTGGCAGCCC AGCTGGAGGC

              451                                                 500
SPLICE FORM 2 ACCCCGGCCT CCCCGGCTCC GCTACCTGCT GGTAGTTTCT ACACGAGAAG
SPLICE FORM 1 ACCCCGGCCT CCCCGGCTCC GCTACCTGCT GGTAGTTTCT ACACGAGAAG
SPLICE FORM 3 ACCCCGGCCT CCCCGGCTCC GCTACCTGCT GGTAGTTTCT ACACGAGAAG

              501                                                 550
SPLICE FORM 2 GAGAAGGTCT GAGCCAGGAT GAGACGGTCC TCCTGGGCGT GGATTTCCCT
SPLICE FORM 1 GAGAAGGTCT GAGCCAGGAT GAGACGGTCC TCCTGGGCGT GGATTTCCCT
SPLICE FORM 3 GAGAAGGTCT GAGCCAGGAT GAGACGGTCC TCCTGGGCGT GGATTTCCCT

              551                                                 600
SPLICE FORM 2 GACAGCAGCT CCCCCAGCTG CACCCTGGGC CTGGTCTTGC CCCTCTGGAG
SPLICE FORM 1 GACAGCAGCT CCCCCAGCTG CACCCTGGGC CTGGTCTTGC CCCTCTGGAG
SPLICE FORM 3 GACAGCAGCT CCCCCAGCTG CACCCTGGGC CTGGTCTTGC CCCTCTGGAG

              601                                                 650
SPLICE FORM 2 TGACACCCAG GTGTACTTAG ATGGAGACGG GGGCTTCAGC GTGACGTCTG
SPLICE FORM 1 TGACACCCAG GTGTACTTAG ATGGAGACGG GGGCTTCAGC GTGACGTCTG
SPLICE FORM 3 TGACACCCAG GTGTACTTAG ATGGAGACGG GGGCTTCAGC GTGACGTCTG

              651                                                 700
SPLICE FORM 2 GTGGGCAAAG CCGGATCTTC AAGCCCATCT CCATCCAGAC CATGTGGGCC
SPLICE FORM 1 GTGGGCAAAG CCGGATCTTC AAGCCCATCT CCATCCAGAC CATGTGGGCC
SPLICE FORM 3 GTGGGCAAAG CCGGATCTTC AAGCCCATCT CCATCCAGAC CATGT.....

              701                                                 750
SPLICE FORM 2 ACACTCCAGG TATTGCACCA AGCATGTGAG GCAGCTCTAG GCAGCGGCCT
SPLICE FORM 1 ACACTCCAGG TATTGCACCA AGCATGTGAG GCAGCTCTAG GCAGCGGCCT
SPLICE FORM 3 .......... .......... .......... .......... ..........

              751                                                 800
SPLICE FORM 2 TGTACCGGGT GGCAGTGCCC TCACCTGGGC CAGCCACTAC CAGGAGAGAC
SPLICE FORM 1 TGTACCGGGT GGCAGTGCCC TCACCTGGGC CAGCCACTAC CAGGAGAGAC
SPLICE FORM 3 .......... .......... .......... .......... ..........
```

FIGURE 3C

```
                801                                                 850
SPLICE FORM 2 TGAACTCCGA ACAGAGCTGC CTCAATGAGT GGACGGCTAT GGCCGACCTG
SPLICE FORM 1 TGAACTCCGA ACAGAGCTGC CTCAATGAGT GGACGGCTAT GGCCGACCTG
SPLICE FORM 3 .......... .......... .......... .......... ..........

                851                                                 900
SPLICE FORM 2 GAGTCTCTGC GGCCTCCCAG CGCCGAGCCT GGCGGGTCCT CAGAACAGGA
SPLICE FORM 1 GAGTCTCTGC GGCCTCCCAG CGCCGAGCCT GGCGGGTCCT CAGAACAGGA
SPLICE FORM 3 .......... .......... .......... ....GGTCCT CAGAACAGGA

                901                                                 950
SPLICE FORM 2 GCAGATGGAG CAGGCGATCC GTGCTGAGCT GTGGAAAGTG TTGGATGTCA
SPLICE FORM 1 GCAGATGGAG CAGGCGATCC GTGCTGAGCT GTGGAAAGTG TTGGATGTCA
SPLICE FORM 3 GCAGATGGAG CAGGCGATCC GTGCTGAGCT GTGGAAAGTG TTGGATGTCA

                951                                                1000
SPLICE FORM 2 GTGACCTGGA GAGTGTCACT TCCAAAGAGA TCCGCCAGGC TCTGGAGCTG
SPLICE FORM 1 GTGACCTGGA GAGTGTCACT TCCAAAGAGA TCCGCCAGGC TCTGGAGCTG
SPLICE FORM 3 GTGACCTGGA GAGTGTCACT TCCAAAGAGA TCCGCCAGGC TCTGGAGCTG

               1001                                                1050
SPLICE FORM 2 CGCCTGGGGC TCCCCCTCCA GCAGTACCGT GACTTCATCG ACAACCAGAT
SPLICE FORM 1 CGCCTGGGGC TCCCCCTCCA GCAGTACCGT GACTTCATCG ACAACCAGAT
SPLICE FORM 3 CGCCTGGGGC TCCCCCTCCA GCAGTACCGT GACTTCATCG ACAACCAGAT

               1051                                                1100
SPLICE FORM 2 GCTGCTGCTG GTGGCACAGC GGGACCGAGC CTCCCGCATC TTCCCCCACC
SPLICE FORM 1 GCTGCTGCTG GTGGCACAGC GGGACCGAGC CTCCCGCATC TTCCCCCACC
SPLICE FORM 3 GCTGCTGCTG GTGGCACAGC GGGACCGAGC CTCCCGCATC TTCCCCCACC

               1101                                                1150
SPLICE FORM 2 TCTACCTGGG CTCAGAGTGG AACGCAGCAA ACCTGGAGGA GCTGCAGAGG
SPLICE FORM 1 TCTACCTGGG CTCAGAGTGG AACGCAGCAA ACCTGGAGGA GCTGCAGAGG
SPLICE FORM 3 TCTACCTGGG CTCAGAGTGG AACGCAGCAA ACCTGGAGGA GCTGCAGAGG

               1151                                                1200
SPLICE FORM 2 AACAGGGTCA CCCACATCTT GAACATGGCC CGGGAGATTG ACAACTTCTA
SPLICE FORM 1 AACAGGGTCA CCCACATCTT GAACATGGCC CGGGAGATTG ACAACTTCTA
SPLICE FORM 3 AACAGGGTCA CCCACATCTT GAACATGGCC CGGGAGATTG ACAACTTCTA

               1201                                                1250
SPLICE FORM 2 CCCTGAGCGC TTCACCTACC ACAATGTGCG CCTCTGGGAT GAGGAGTCGG
SPLICE FORM 1 CCCTGAGCGC TTCACCTACC ACAATGTGCG CCTCTGGGAT GAGGAGTCGG
SPLICE FORM 3 CCCTGAGCGC TTCACCTACC ACAATGTGCG CCTCTGGGAT GAGGAGTCGG

               1251                                                1300
SPLICE FORM 2 CCCAGCTGCT GCCGCACTGG AAGGAGACGC ACCGCTTCAT TGAGGCTGCA
SPLICE FORM 1 CCCAGCTGCT GCCGCACTGG AAGGAGACGC ACCGCTTCAT TGAGGCTGCA
SPLICE FORM 3 CCCAGCTGCT GCCGCACTGG AAGGAGACGC ACCGCTTCAT TGAGGCTGCA

               1301                                                1350
SPLICE FORM 2 AGAGCACAGG GCACCCACGT GCTGGTCCAC TGCAAGATGG GCGTCAGCCG
SPLICE FORM 1 AGAGCACAGG GCACCCACGT GCTGGTCCAC TGCAAGATGG GCGTCAGCCG
SPLICE FORM 3 AGAGCACAGG GCACCCACGT GCTGGTCCAC TGCAAGATGG GCGTCAGCCG

               1351                                                1400
SPLICE FORM 2 CTCAGCGGCC ACAGTGCTGG CCTATGCCAT GAAGCAGTAC GAATGCAGCC
SPLICE FORM 1 CTCAGCGGCC ACAGTGCTGG CCTATGCCAT GAAGCAGTAC GAATGCAGCC
SPLICE FORM 3 CTCAGCGGCC ACAGTGCTGG CCTATGCCAT GAAGCAGTAC GAATGCAGCC

               1401                                                1450
SPLICE FORM 2 TGGAGCAGGC CCTGCGCCAC GTGCAGGAGC TCCGGCCCAT CGCCCGCCCC
SPLICE FORM 1 TGGAGCAGGC CCTGCGCCAC GTGCAGGAGC TCCGGCCCAT CGCCCGCCCC
SPLICE FORM 3 TGGAGCAGGC CCTGCGCCAC GTGCAGGAGC TCCGGCCCAT CGCCCGCCCC
```

FIGURE 3D

```
                1451                                                1500
SPLICE FORM 2 AACCCTGGCT TCCTGCGCCA GCTGCAGATC TACCAGGGCA TCCTGACGGC
SPLICE FORM 1 AACCCTGGCT TCCTGCGCCA GCTGCAGATC TACCAGGGCA TCCTGACG..
SPLICE FORM 3 AACCCTGGCT TCCTGCGCCA GCTGCAGATC TACCAGGGCA TCCTGACG..

                1501                                                1550
SPLICE FORM 2 CAGCCGCCAG AGCCATGTCT GGGAGCAGAA AGTGGGTGGG GTCTCCCCAG
SPLICE FORM 1 .......... .......... .......... .......... ..........
SPLICE FORM 3 .......... .......... .......... .......... ..........

                1551                                                1600
SPLICE FORM 2 AGGAGCACCC AGCCCCTGAA GTCTCTACAC CATTCCCACT TCTTCCGCCA
SPLICE FORM 1 .......... .......... .......... .......... ......GCCA
SPLICE FORM 3 .......... .......... .......... .......... ......GCCA

                1601                                                1650
SPLICE FORM 2 GAACCTGAGG GTGGTGGGGA GGAGAAGGTT GTAGGCATGG AAGAGAGCCA
SPLICE FORM 1 GAACCTGAGG GTGGTGGGGA GGAGAAGGTT GTAGGCATGG AAGAGAGCCA
SPLICE FORM 3 GAACCTGAGG GTGGTGGGGA GGAGAAGGTT GTAGGCATGG AAGAGAGCCA

                1651                                                1700
SPLICE FORM 2 GGCAGCCCCG AAAGAAGAGC CTGGG..CCA CGGCCACGTA TAAACCTCCG
SPLICE FORM 1 GGCAGCCCCG AAAGAAGAGC CTGGG..CCA CGGCCACGTA TAAACCTCCG
SPLICE FORM 3 GGCAGCCCCG AAAGAAGAGC CTGGGGCCAC GGGGCACGTA TAAACCTCCG

                1701                                                1750
SPLICE FORM 2 AGGGGTCATG AGGTCCATCA GTCTTCTGGA GCCCTCCTT. GGAGCTGGAG
SPLICE FORM 1 AGGGGTCATG AGGTCCATCA GTCTTCTGGA GCCCTCCTT. GGAGCTGGAG
SPLICE FORM 3 AGGGGTCATG AGGTCCATCA GTCTTCTGGA GCCCTCCTTG GGAGCTGGAG

                1751                                                1800
SPLICE FORM 2 AGCACCTCAG .AGACCAGTG ACATGCCAGA GGTCTTCTCT TCCCACGAGT
SPLICE FORM 1 AGCACCTCAG .AGACCAGTG ACATGCCAGA GGTCTTCTCT TCCCACGAGT
SPLICE FORM 3 AGCACCTCAG TAGACCAGTG ACATGCCAGA GGTCTTCTCT TCCCACGAGT

                1801                                                1850
SPLICE FORM 2 CTTCACATGA AGAGCCTCTG CAGCCCTTCC CACAGCTTGC AAGGACCAAG
SPLICE FORM 1 CTTCACATGA AGAGCCTCTG CAGCCCTTCC CACAGCTTGC AAGGACCAAG
SPLICE FORM 3 CTTCACATGA AGAGCCTCTG CAGCCCTTCC CACAGCTTGC AAGGACCAAG

                1851                                                1900
SPLICE FORM 2 GGAGGCCAGC AGGTGGACAG GGGGCCTCAG CCTGCCCTGA AGTCCCGCCA
SPLICE FORM 1 GGAGGCCAGC AGGTGGACAG GGGGCCTCAG CCTGCCCTGA AGTCCCGCCA
SPLICE FORM 3 GGAGGCCAGC AGGTGGACAG GGGGCCTCAG CCTGCCCTGA AGTCCCGCCA

                1901                                                1950
SPLICE FORM 2 GTCAGTGGTT ACCCTCCAGG GCAGTGCCGT GGTGGCCAAC CGGACCCAGG
SPLICE FORM 1 GTCAGTGGTT ACCCTCCAGG GCAGTGCCGT GGTGGCCAAC CGGACCCAGG
SPLICE FORM 3 GTCAGTGGTT ACCCTCCAGG GCAGTGCCGT GGTGGCCAAC CGGACCCAGG

                1951                                                2000
SPLICE FORM 2 CCTTCCAGGA GCAGGAGCAG GGGCAGGGGC AGGGGCAGGG AGAGCCCTGC
SPLICE FORM 1 CCTTCCAGGA GCAGGAGCAG GGGCAGGGGC AGGGGCAGGG AGAGCCCTGC
SPLICE FORM 3 CCTTCCAGGA GCAGGAGCAG GGGCAGGGGC AGGGGCAGGG AGAGCCCTGC

                2001                                                2050
SPLICE FORM 2 ATTTCCTCTA CGCCCAGGTT CCGGAAGGTG GTGAGACAGG CCAGCGTGCA
SPLICE FORM 1 ATTTCCTCTA CGCCCAGGTT CCGGAAGGTG GTGAGACAGG CCAGCGTGCA
SPLICE FORM 3 ATTTCCTCTA CGCCCAGGTT CCGGAAGGTG GTGAGACAGG CCAGCGTGCA

                2051                                                2100
SPLICE FORM 2 TGACAGTGGA GAGGAGGGCG AGGCCTGAGC CCTCACACAT GCCCACGCTC
SPLICE FORM 1 TGACAGTGGA GAGGAGGGCG AGGCCTGAGC CCTCACACAT GCCCACGCTC
SPLICE FORM 3 TGACAGTGGA GAGGAGGGCG AGGCCTGAGC CCTCACACAT GCCCACGCTC

                2101                                                2150
```

FIGURE 3E

```
SPLICE FORM 2 CCCTGACACT GAAGAGGATC CACAACTCCT TGGAGAAACA CCCTCACGTC
SPLICE FORM 1 CCCTGACACT GAAGAGGATC CACAACTCCT TGGAGAAACA CCCTCACGTC
SPLICE FORM 3 CCCTGACACT GAAGAGGATC CACAACTCCT TGGAGAAACA CCCTCACGTC

              2151                                              2200
SPLICE FORM 2 TGTTGCCGCA CACATTCCTC TCAGCTCCGC CCCATACCCG TCACTACAGC
SPLICE FORM 1 TGTTGCCGCA CACATTCCTC TCAGCTCCGC CCCATACCCG TCACTACAGC
SPLICE FORM 3 TGTTGCCGCA CACATTCCTC TCAGCTCCGC CCCATACCCG TCACTACAGC

              2201                                              2250
SPLICE FORM 2 CTCACCTCCC ACCCCTGTCA CTACGGCCTC ACCTCCCACC CCTGTCACTA
SPLICE FORM 1 CTCACCTCCC ACCCCTGTCA CTACGGCCTC ACCTCCCACC CCTGTCACTA
SPLICE FORM 3 CTCACCTCCC ACCCCTGTCA CTACGGCCTC ACCTCCCACC CCTGTCACTA

              2251                                              2300
SPLICE FORM 2 CAGCCTCACC TCCTACAGCC TTAAGTCCCA GGCCCATGTC TGCCTGTCCA
SPLICE FORM 1 CAGCCTCACC TCCTACAGCC TTAAGTCCCA GGCCCATGTC TGCCTGTCCA
SPLICE FORM 3 CAGCCTCACC TCCTACAGCC TTAAGTCCCA GGCCCATGTC TGCCTGTCCA

              2301                                              2350
SPLICE FORM 2 AGGGCTCAAG ACTTTCTAAC TGGGATGTGG TAGAGGGACT GAAGGTACCT
SPLICE FORM 1 AGGGCTCAAG ACTTTCTAAC TGGGATGTGG TAGAGGGACT GAAGGTACCT
SPLICE FORM 3 AGGGCTCAAG ACTTTCTAAC TGGGATGTGG TAGAGGGACT GAAGGTACCT

              2351                                              2400
SPLICE FORM 2 TTGGGGGCAA CAGCACCCTA GTTTCATTCT CAACTCTAGC CCTGCACACT
SPLICE FORM 1 TTGGGGGCAA CAGCACCCTA GTTTCATTCT CAACTCTAGC CCTGCACACT
SPLICE FORM 3 TTGGGGGCAA CAGCACCCTA GTTTCATTCT CAACTCTAGC CCTGCACACT

              2401                                              2450
SPLICE FORM 2 CACCTGTGGC ACGGAATGAA AACAGAGCTT CCCGTGCAAA AAGGGTCACG
SPLICE FORM 1 CACCTGTGGC ACGGAATGAA AACAGAGCTT CCCGTGCAAA AAGGGTCACG
SPLICE FORM 3 CACCTGTGGC ACGGAATGAA AACAGAGCTT CCCGTGCAAA AAGGGTCACG

              2451                                              2500
SPLICE FORM 2 CCTCCCACCC CCGCCCCCTC CCTGCACCTC CTGTCCTCTC CCAGTTCATT
SPLICE FORM 1 CCTCCCACCC CCGCCCCCTC CCTGCACCTC CTGTCCTCTC CCAGTTCATT
SPLICE FORM 3 CCTCCCACCC CCGCCCCCTC CCTGCACCTC CTGTCCTCTC CCAGTTCATT

              2501                                              2550
SPLICE FORM 2 CCTGGAACCA GCCAGGCCAG GCAACCAGTG GCCCCCAAAG GCAGGCAGGA
SPLICE FORM 1 CCTGGAACCA GCCAGGCCAG GCAACCAGTG GCCCCCAAAG GCAGGCAGGA
SPLICE FORM 3 CCTGGAACCA GCCAGGCCAG GCAACCAGTG GCCCCCAAAG GCAGGCAGGA

              2551                                              2600
SPLICE FORM 2 TCCTCAGGCC CCAGCCGCGG GAGGCTGGAA GGGCTGGCAG ATCGCTTCCC
SPLICE FORM 1 TCCTCAGGCC CCAGCCGCGG GAGGCTGGAA GGGCTGGCAG ATCGCTTCCC
SPLICE FORM 3 TCCTCAGGCC CCAGCCGCGG GAGGCTGGAA GGGCTGGCAG ATCGCTTCCC

              2601                                              2650
SPLICE FORM 2 TCATCCACCT CCACCGGTCC AGGTCTTTGC TGCTGTCCCC AGACCTCCTG
SPLICE FORM 1 TCATCCACCT CCACCGGTCC AGGTCTTTGC TGCTGTCCCC AGACCTCCTG
SPLICE FORM 3 TCATCCACCT CCACCGGTCC AGGTCTTTGC TGCTGTCCCC AGACCTCCTG

              2651                                              2700
SPLICE FORM 2 TGACACCACG CCAGATCACA GGGCACCAGG CCAGAGATAG TCTTCTTTTT
SPLICE FORM 1 TGACACCACG CCAGATCACA GGGCACCAGG CCAGAGATAG TCTTCTTTTT
SPLICE FORM 3 TGACACCACG CCAGATCACA GGGCACCAGG CCAGAGATAG TCTTCTTTTT

              2701                                              2750
SPLICE FORM 2 GTCCTTTCTG GCCTCTGGCT AGTCAGTTTT TCATAGCCTT ACAGTATCTG
SPLICE FORM 1 GTCCTTTCTG GCCTCTGGCT AGTCAGTTTT TCATAGCCTT ACAGTATCTG
SPLICE FORM 3 GTCCTTTCTG GCCTCTGGCT AGTCAGTTTT TCATAGCCTT ACAGTATCTG

              2751                                              2800
SPLICE FORM 2 GCTTTGTACT GAGAAATAAA ACACATTTTC ATAAAAAAAA AAAAAAAAAA
```

FIGURE 3F

```
SPLICE FORM 1 GCTTTGTACT GAGAAATAAA ACACATTTTC ATAAAAAAAA AAAAAAAAAA
SPLICE FORM 3 GCTTTGTACT GAGAAATAAA ACACATTTTC ATATTTGGTT AAAAAAAAAA

              2801                                                2850
SPLICE FORM 2 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA
SPLICE FORM 1 AAAAAAA~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SPLICE FORM 3 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA~~~~~~ ~~~~~~~~~~

              2851                                          2894
SPLICE FORM 2 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA
SPLICE FORM 1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~
SPLICE FORM 3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~
```

MULTIPLE ALIGNMENT OF PEPTIDE SEQUENCES:

```
SPLICE FORM 2 MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD
SPLICE FORM 1 MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD
SPLICE FORM 3 MALVTVSRSP PGSGASTPVG PWDQAVQRRS RLQRRQSFAV LRGAVLGLQD

              51                                                  100
SPLICE FORM 2 GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL
SPLICE FORM 1 GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL
SPLICE FORM 3 GGDNDDAAEA SSEPTEKAPS EEELHGDQTD FGQGSQSPQK QEEQRQHLHL

              101                                                 150
SPLICE FORM 2 MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD
SPLICE FORM 1 MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD
SPLICE FORM 3 MVQLLRPQDD IRLAAQLEAP RPPRLRYLLV VSTREGEGLS QDETVLLGVD

              151                                                 200
SPLICE FORM 2 FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM
SPLICE FORM 1 FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM
SPLICE FORM 3 FPDSSSPSCT LGLVLPLWSD TQVYLDGDGG FSVTSGGQSR IFKPISIQTM

              201                                                 250
SPLICE FORM 2 WATLQVLHQA CEAALGSGLV PGGSALTWAS HYQERLNSEQ SCLNEWTAMA
SPLICE FORM 1 WATLQVLHQA CEAALGSGLV PGGSALTWAS HYQERLNSEQ SCLNEWTAMA
SPLICE FORM 3 W..........  ..........  ..........  ..........  ..........

              251                                                 300
SPLICE FORM 2 DLESLRPPSA EPGGSSEQEQ MEQAIRAELW KVLDVSDLES VTSKEIRQAL
SPLICE FORM 1 DLESLRPPSA EPGGSSEQEQ MEQAIRAELW KVLDVSDLES VTSKEIRQAL
SPLICE FORM 3 ..........  ....SSEQEQ MEQAIRAELW KVLDVSDLES VTSKEIRQAL

              301                                                 350
SPLICE FORM 2 ELRLGLPLQQ YRDFIDNQML LLVAQRDRAS RIFPHLYLGS EWNAANLEEL
SPLICE FORM 1 ELRLGLPLQQ YRDFIDNQML LLVAQRDRAS RIFPHLYLGS EWNAANLEEL
SPLICE FORM 3 ELRLGLPLQQ YRDFIDNQML LLVAQRDRAS RIFPHLYLGS EWNAANLEEL

              351                                                 400
SPLICE FORM 2 QRNRVTHILN MAREIDNFYP ERFTYHNVRL WDEESAQLLP HWKETHRFIE
SPLICE FORM 1 QRNRVTHILN MAREIDNFYP ERFTYHNVRL WDEESAQLLP HWKETHRFIE
SPLICE FORM 3 QRNRVTHILN MAREIDNFYP ERFTYHNVRL WDEESAQLLP HWKETHRFIE

              401                                                 450
SPLICE FORM 2 AARAQGTHVL VHCKMGVSRS AATVLAYAMK QYECSLEQAL RHVQELRPIA
SPLICE FORM 1 AARAQGTHVL VHCKMGVSRS AATVLAYAMK QYECSLEQAL RHVQELRPIA
SPLICE FORM 3 AARAQGTHVL VHCKMGVSRS AATVLAYAMK QYECSLEQAL RHVQELRPIA

              451                                                 500
SPLICE FORM 2 RPNPGFLRQL QIYQGILTAS RQSHVWEQKV GGVSPEEHPA PEVSTPFPLL
SPLICE FORM 1 RPNPGFLRQL QIYQGILTAR T~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SPLICE FORM 3 RPNPGFLRQL QIYQGILTAR T~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              501                                                 550
SPLICE FORM 2 PPEPEGGGEE KVVGMEESQA APKEEPGPRP RINLRGVMRS ISLLEPSLEL
```

FIGURE 3G

```
SPLICE FORM 1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SPLICE FORM 3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              551                                                  600
SPLICE FORM 2 ESTSETSDMP EVFSSHESSH EEPLQPFPQL ARTKGGQQVD RGPQPALKSR
SPLICE FORM 1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SPLICE FORM 3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              601                                                  650
SPLICE FORM 2 QSVVTLQGSA VVANRTQAFQ EQEQGQGQGQ GEPCISSTPR FRKVVRQASV
SPLICE FORM 1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
SPLICE FORM 3 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

              651
SPLICE FORM 2 HDSGEEGEA
SPLICE FORM 1 ~~~~~~~~~
SPLICE FORM 3 ~~~~~~~~~
```

FIGURE 3H

SEQUENCE LISTING

<110> PE CORPORATION (NY)

<120> ISOLATED HUMAN PHOSPHATASE PROTEINS,
NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
AND USES THEREOF

<130> CL000964PCT

<150> TO BE ASSIGNED
<151> 2001-11-07

<140> 09/761,640
<141> 2001-01-18

<140> 09/715,177
<141> 2000-11-20

<160> 10

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 2704
<212> DNA
<213> Homo sapiens

<400> 1
cgtccttcct ggtcctgcgg gtccaggact gtccgcgggg ttgagggaag
gggccgtgcc 60
cggtgccagc ccaggtgctc gcggcctggc tccatggccc tggtcacagt
gagccgttcg 120
cccccgggca gcggcgcctc cacgcccgtg gggccctggg accaggcggt
ccagcgaagg 180
agtcgactcc agcgaaggca gagctttgcg gtgctccgtg gggctgtcct
gggactgcag 240
gatggagggg acaatgatga tgcagcagag gccagttctg agccaacaga
gaaggccccg 300
agtgaggagg agctccacgg ggaccagaca gacttcgggc aaggatccca
gagtccccag 360
aagcaggagg agcagaggca gcacctgcac ctcatggtac agctgctgag
gccgcaggat 420
gacatccgcc tggcagccca gctggaggca ccccggcctc cccggctccg
ctacctgctg 480
gtagtttcta cacgagaagg agaaggtctg agccaggatg agacggtcct
cctgggcgtg 540
gatttccctg acagcagctc ccccagctgc accctgggcc tggtcttgcc
cctctggagt 600
gacacccagg tgtacttaga tggagacggg ggcttcagcg tgacgtctgg
tgggcaaagc 660
cggatcttca agcccatctc catccagacc atgtgggcca cactccaggt
attgcaccaa 720
gcatgtgagg cagctctagg cagcggcctt gtaccgggtg gcagtgccct
cacctgggcc 780
agccactacc aggagagact gaactccgaa cagagctgcc tcaatgagtg
gacggctatg 840
gccgacctgg agtctctgcg gcctcccagc gccgagcctg gcgggtcctc
agaacaggag 900

cagatggagc aggcgatccg tgctgagctg tggaaagtgt tggatgtcag
tgacctggag 960
agtgtcactt ccaaagagat ccgccaggct ctggagctgc gcctggggct
ccccctccag 1020
cagtaccgtg acttcatcga caaccagatg ctgctgctgg tggcacagcg
ggaccgagcc 1080
tcccgcatct tcccccacct ctacctgggc tcagagtgga acgcagcaaa
cctggaggag 1140
ctgcagagga acagggtcac ccacatcttg aacatggccc gggagattga
caacttctac 1200
cctgagcgct tcacctacca caatgtgcgc ctctgggatg aggagtcggc
ccagctgctg 1260
ccgcactgga aggagacgca ccgcttcatt gaggctgcaa gagcacaggg
cacccacgtg 1320
ctggtccact gcaagatggg cgtcagccgc tcagcggcca cagtgctggc
ctatgccatg 1380
aagcagtacg aatgcagcct ggagcaggcc ctgcgccacg tgcaggagct
ccggcccatc 1440
gcccgcccca accctggctt cctgcgccag ctgcagatct accagggcat
cctgacggcc 1500
agaacctgag ggtggtgggg aggagaaggt tgtaggcatg gaagagagcc
aggcagcccc 1560
gaaagaagag cctgggccac ggccacgtat aaacctccga ggggtcatga
ggtccatcag 1620
tcttctggag ccctccttgg agctggagag cacctcagag accagtgaca
tgccagaggt 1680
cttctcttcc cacgagtctt cacatgaaga gcctctgcag cccttcccac
agcttgcaag 1740
gaccaaggga ggccagcagg tggacagggg gcctcagcct gccctgaagt
cccgccagtc 1800
agtggttacc ctccagggca gtgccgtggt ggccaaccgg acccaggcct
tccaggagca 1860
ggagcagggg caggggcagg ggcagggaga gccctgcatt tcctctacgc
ccaggttccg 1920
gaaggtggtg agacaggcca gcgtgcatga cagtggagag gagggcgagg
cctgagccct 1980
cacacatgcc cacgctcccc tgacactgaa gaggatccac aactccttgg
agaaacaccc 2040
tcacgtctgt tgccgcacac attcctctca gctccgcccc atacccgtca
ctacagcctc 2100
acctcccacc cctgtcacta cggcctcacc tcccaccccc gtcactacag
cctcacctcc 2160
tacagcctta agtcccaggc ccatgtctgc ctgtccaagg gctcaagact
ttctaactgg 2220
gatgtggtag agggactgaa ggtacctttg ggggcaacag caccctagtt
tcattctcaa 2280
ctctagccct gcacactcac ctgtggcacg gaatgaaaac agagcttccc
gtgcaaaaag 2340
ggtcacgcct cccacccccg ccccctccct gcacctcctg tcctctccca
gttcattcct 2400
ggaaccagcc aggccaggca accagtggcc cccaaaggca ggcaggatcc
tcaggcccca 2460
gccgcgggag gctggaaggg ctggcagatc gcttccctca tccacctcca
ccggtccagg 2520
tctttgctgc tgtccccaga cctcctgtga caccacgcca gatcacaggg
caccaggcca 2580
gagatagtct tctttttgtc ctttctggcc tctggctagt cagtttttca
tagccttaca 2640
gtatctggct ttgtactgag aaataaaaca cattttcata aaaaaaaaaa
aaaaaaaaaa 2700

aaaa
2704

<210> 2
<211> 2852
<212> DNA
<213> Homo sapiens

<400> 2
tggttgaggg aaggggccgt gcccggtgcc agcccaggtg ctcgcggcct
ggctccatgg 60
ccctggtcac agtgagccgt tcgcccccgg gcagcggcgc ctccacgccc
gtggggccct 120
gggaccaggc ggtccagcga aggagtcgac tccagcgaag gcagagcttt
gcggtgctcc 180
gtggggctgt cctgggactg caggatggag gggacaatga tgatgcagca
gaggccagtt 240
ctgagccaac agagaaggcc ccgagtgagg aggagctcca cggggaccag
acagacttcg 300
ggcaaggatc ccagagtccc cagaagcagg aggagcagag gcagcacctg
cacctcatgg 360
tacagctgct gaggccgcag gatgacatcc gcctggcagc ccagctggag
gcaccccggc 420
ctccccggct ccgctacctg ctggtagttt ctacacgaga aggagaaggt
ctgagccagg 480
atgagacggt cctcctgggc gtggatttcc ctgacagcag ctcccccagc
tgcaccctgg 540
gcctggtctt gcccctctgg agtgacaccc aggtgtactt agatggagac
gggggcttca 600
gcgtgacgtc tggtgggcaa agccggatct tcaagcccat ctccatccag
accatgtggg 660
ccacactcca ggtattgcac caagcatgtg aggcagctct aggcagcggc
cttgtaccgg 720
gtggcagtgc cctcacctgg gccagccact accaggagag actgaactcc
gaacagagct 780
gcctcaatga gtggacggct atggccgacc tggagtctct gcggcctccc
agcgccgagc 840
ctggcgggtc ctcagaacag gagcagatgg agcaggcgat ccgtgctgag
ctgtggaaag 900
tgttggatgt cagtgacctg gagagtgtca cttccaaaga gatccgccag
gctctggagc 960
tgcgcctggg gctccccctc cagcagtacc gtgacttcat cgacaaccag
atgctgctgc 1020
tggtggcaca gcgggaccga gcctcccgca tcttccccca cctctacctg
ggctcagagt 1080
ggaacgcagc aaacctggag gagctgcaga ggaacagggt cacccacatc
ttgaacatgg 1140
cccgggagat tgacaacttc taccctgagc gcttcaccta ccacaatgtg
cgcctctggg 1200
atgaggagtc ggcccagctg ctgccgcact ggaaggagac gcaccgcttc
attgaggctg 1260
caagagcaca gggcacccac gtgctggtcc actgcaagat gggcgtcagc
cgctcagcgg 1320
ccacagtgct ggcctatgcc atgaagcagt acgaatgcag cctggagcag
gccctgcgcc 1380
acgtgcagga gctccggccc atcgcccgcc ccaaccctgg cttcctgcgc
cagctgcaga 1440
tctaccaggg catcctgacg gccagccgcc agagccatgt ctgggagcag
aaagtgggtg 1500
gggtctcccc agaggagcac ccagcccctg aagtctctac accattccca
cttcttccgc 1560

cagaacctga gggtggtggg gaggagaagg ttgtaggcat ggaagagagc
caggcagccc 1620
cgaaagaaga gcctgggcca cggccacgta taaacctccg aggggtcatg
aggtccatca 1680
gtcttctgga gccctccttg gagctggaga gcacctcaga gaccagtgac
atgccagagg 1740
tcttctcttc ccacgagtct tcacatgaag agcctctgca gcccttccca
cagcttgcaa 1800
ggaccaaggg aggccagcag gtggacaggg ggcctcagcc tgccctgaag
tcccgccagt 1860
cagtggttac cctccagggc agtgccgtgg tggccaaccg gacccaggcc
ttccaggagc 1920
aggagcaggg gcaggggcag gggcagggag agccctgcat ttcctctacg
cccaggttcc 1980
ggaaggtggt gagacaggcc agcgtgcatg acagtggaga ggagggcgag
gcctgagccc 2040
tcacacatgc ccacgctccc ctgacactga agaggatcca caactccttg
gagaaacacc 2100
ctcacgtctg ttgccgcaca cattcctctc agctccgccc catacccgtc
actacagcct 2160
cacctcccac ccctgtcact acggcctcac ctcccacccc tgtcactaca
gcctcacctc 2220
ctacagcctt aagtcccagg cccatgtctg cctgtccaag ggctcaagac
tttctaactg 2280
ggatgtggta gagggactga aggtaccttt gggggcaaca gcaccctagt
ttcattctca 2340
actctagccc tgcacactca cctgtggcac ggaatgaaaa cagagcttcc
cgtgcaaaaa 2400
gggtcacgcc tcccacccca gccccctccc tgcacctcct gtcctctccc
agttcattcc 2460
tggaaccagc caggccaggc aaccagtggc ccccaaaggc aggcaggatc
ctcaggcccc 2520
agccgcggga ggctggaagg gctggcagat cgcttccctc atccacctcc
accggtccag 2580
gtctttgctg ctgtccccag acctcctgtg acaccacgcc agatcacagg
gcaccaggcc 2640
agagatagtc ttctttttgt cctttctggc ctctggctag tcagtttttc
atagccttac 2700
agtatctggc tttgtactga gaaataaaac acattttcat aaaaaaaaaa
aaaaaaaaaa 2760
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa
aaaaaaaaaa 2820
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa
2852

<210> 3
<211> 2540
<212> DNA
<213> Homo sapiens

<400> 3
cctggtcctg cgggtccagg actgtcccgc ggggttgagg gaaggggccg
tgcccggtgc 60
cagcccaggt gctcgcggcc tggctccatg gccctggtca cagtgagccg
ttcgcccccg 120
ggcagcggcg cctccacgcc cgtggggccc tgggaccagg cggtccagcg
aaggagtcga 180
ctccagcgaa ggcagagctt tgcggtgctc cgtggggctg tcctgggact
gcaggatgga 240
ggggacaatg atgatgcagc agaggccagt tctgagccaa cagagaaggc
cgccgagtgag 300

gaggagctcc acggggacca gacagacttc gggcaaggat cccagagtcc
ccagaagcag 360
gaggagcaga ggcagcacct gcacctcatg gtacagctgc tgaggccgca
ggatgacatc 420
cgcctggcag cccagctgga ggcaccccgg cctccccggc tccgctacct
gctggtagtt 480
tctacacgag aaggagaagg tctgagccag gatgagacgg tcctcctggg
cgtggatttc 540
cctgacagca gctcccccag ctgcaccctg ggcctggtct tgcccctctg
gagtgacacc 600
caggtgtact tagatggaga cgggggcttc agcgtgacgt ctggtgggca
aagccggatc 660
ttcaagccca tctccatcca gaccatgtgg tcctcagaac aggagcagat
ggagcaggcg 720
atccgtgctg agctgtggaa agtgttggat gtcagtgacc tggagagtgt
cacttccaaa 780
gagatccgcc aggctctgga gctgcgcctg gggctccccc tccagcagta
ccgtgacttc 840
atcgacaacc agatgctgct gctggtggca cagcgggacc gagcctcccg
catcttcccc 900
cacctctacc tgggctcaga gtggaacgca gcaaacctgg aggagctgca
gaggaacagg 960
gtcacccaca tcttgaacat ggcccgggag attgacaact tctaccctga
gcgcttcacc 1020
taccacaatg tgcgcctctg ggatgaggag tcggcccagc tgctgccgca
ctggaaggag 1080
acgcaccgct tcattgaggc tgcaagagca cagggcaccc acgtgctggt
ccactgcaag 1140
atgggcgtca gccgctcagc ggccacagtg ctggcctatg ccatgaagca
gtacgaatgc 1200
agcctggagc aggccctgcg ccacgtgcag gagctccggc ccatcgcccg
ccccaaccct 1260
ggcttcctgc gccagctgca gatctaccag ggcatcctga cggccagaac
ctgagggtgg 1320
tggggaggag aaggttgtag gcatggaaga gagccaggca gccccgaaag
aagagcctgg 1380
ggccacgggg cacgtataaa cctccgaggg gtcatgaggt ccatcagtct
tctggagccc 1440
tccttgggag ctggagagca cctcagtaga ccagtgacat gccagaggtc
ttctcttccc 1500
acgagtcttc acatgaagag cctctgcagc ccttcccaca gcttgcaagg
accaagggag 1560
gccagcaggt ggacaggggg cctcagcctg ccctgaagtc ccgccagtca
gtggttaccc 1620
tccagggcag tgccgtggtg gccaaccgga cccaggcctt ccaggagcag
gagcaggggc 1680
aggggcaggg gcagggagag ccctgcattt cctctacgcc caggttccgg
aaggtggtga 1740
gacaggccag cgtgcatgac agtggagagg agggcgaggc ctgagccctc
acacatgccc 1800
acgctcccct gacactgaag aggatccaca actccttgga gaaacaccct
cacgtctgtt 1860
gccgcacaca ttcctctcag ctccgcccca tacccgtcac tacagcctca
cctcccaccc 1920
ctgtcactac ggcctcacct cccacccctg tcactacagc ctcacctcct
acagccttaa 1980
gtcccaggcc catgtctgcc tgtccaaggg ctcaagactt tctaactggg
atgtggtaga 2040
gggactgaag gtacctttgg gggcaacagc accctagttt cattctcaac
tctagccctg 2100

cacactcacc tgtggcacgg aatgaaaaca gagcttcccg tgcaaaaagg
gtcacgcctc 2160
ccacccccgc cccctccctg cacctcctgt cctctcccag ttcattcctg
gaaccagcca 2220
ggccaggcaa ccagtggccc ccaaaggcag gcaggatcct caggccccag
ccgcgggagg 2280
ctggaagggc tggcagatcg cttccctcat ccacctccac cggtccaggt
ctttgctgct 2340
gtccccagac ctcctgtgac accacgccag atcacagggc accaggccag
agatagtctt 2400
ctttttgtcc tttctggcct ctggctagtc agtttttcat agccttacag
tatctggctt 2460
tgtactgaga aataaaacac attttcatat ttggttaaaa aaaaaaaaaa
aaaaaaaaaa 2520
aaaaaaaaaa aaaaaaaaaa
2540

<210> 4
<211> 471
<212> PRT
<213> Homo sapiens

<400> 4
Met Ala Leu Val Thr Val Ser Arg Ser Pro Pro Gly Ser Gly Ala Ser
1 5 10 15
Thr Pro Val Gly Pro Trp Asp Gln Ala Val Gln Arg Arg Ser Arg Leu
20 25 30
Gln Arg Arg Gln Ser Phe Ala Val Leu Arg Gly Ala Val Leu Gly Leu
35 40 45
Gln Asp Gly Gly Asp Asn Asp Asp Ala Ala Glu Ala Ser Ser Glu Pro
50 55 60
Thr Glu Lys Ala Pro Ser Glu Glu Glu Leu His Gly Asp Gln Thr Asp
65 70 75 80
Phe Gly Gln Gly Ser Gln Ser Pro Gln Lys Gln Glu Glu Gln Arg Gln
85 90 95
His Leu His Leu Met Val Gln Leu Leu Arg Pro Gln Asp Asp Ile Arg
100 105 110
Leu Ala Ala Gln Leu Glu Ala Pro Arg Pro Pro Arg Leu Arg Tyr Leu
115 120 125
Leu Val Val Ser Thr Arg Glu Gly Glu Gly Leu Ser Gln Asp Glu Thr
130 135 140
Val Leu Leu Gly Val Asp Phe Pro Asp Ser Ser Ser Pro Ser Cys Thr
145 150 155 160
Leu Gly Leu Val Leu Pro Leu Trp Ser Asp Thr Gln Val Tyr Leu Asp
165 170 175
Gly Asp Gly Gly Phe Ser Val Thr Ser Gly Gly Gln Ser Arg Ile Phe
180 185 190
Lys Pro Ile Ser Ile Gln Thr Met Trp Ala Thr Leu Gln Val Leu His
195 200 205
Gln Ala Cys Glu Ala Ala Leu Gly Ser Gly Leu Val Pro Gly Gly Ser
210 215 220
Ala Leu Thr Trp Ala Ser His Tyr Gln Glu Arg Leu Asn Ser Glu Gln
225 230 235 240
Ser Cys Leu Asn Glu Trp Thr Ala Met Ala Asp Leu Glu Ser Leu Arg
245 250 255
Pro Pro Ser Ala Glu Pro Gly Gly Ser Ser Glu Gln Glu Gln Met Glu
260 265 270
Gln Ala Ile Arg Ala Glu Leu Trp Lys Val Leu Asp Val Ser Asp Leu
275 280 285
Glu Ser Val Thr Ser Lys Glu Ile Arg Gln Ala Leu Glu Leu Arg Leu
290 295 300

```
Gly Leu Pro Leu Gln Gln Tyr Arg Asp Phe Ile Asp Asn Gln Met Leu
305                 310                 315                 320
Leu Leu Val Ala Gln Arg Asp Arg Ala Ser Arg Ile Phe Pro His Leu
                325                 330                 335
Tyr Leu Gly Ser Glu Trp Asn Ala Ala Asn Leu Glu Glu Leu Gln Arg
            340                 345                 350
Asn Arg Val Thr His Ile Leu Asn Met Ala Arg Glu Ile Asp Asn Phe
        355                 360                 365
Tyr Pro Glu Arg Phe Thr Tyr His Asn Val Arg Leu Trp Asp Glu Glu
    370                 375                 380
Ser Ala Gln Leu Leu Pro His Trp Lys Glu Thr His Arg Phe Ile Glu
385                 390                 395                 400
Ala Ala Arg Ala Gln Gly Thr His Val Leu Val His Cys Lys Met Gly
                405                 410                 415
Val Ser Arg Ser Ala Ala Thr Val Leu Ala Tyr Ala Met Lys Gln Tyr
            420                 425                 430
Glu Cys Ser Leu Glu Gln Ala Leu Arg His Val Gln Glu Leu Arg Pro
        435                 440                 445
Ile Ala Arg Pro Asn Pro Gly Phe Leu Arg Gln Leu Gln Ile Tyr Gln
    450                 455                 460
Gly Ile Leu Thr Ala Arg Thr
465                 470
```

<210> 5
<211> 659
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Ala Leu Val Thr Val Ser Arg Ser Pro Pro Gly Ser Gly Ala Ser
 1               5                  10                  15
Thr Pro Val Gly Pro Trp Asp Gln Ala Val Gln Arg Arg Ser Arg Leu
            20                  25                  30
Gln Arg Arg Gln Ser Phe Ala Val Leu Arg Gly Ala Val Leu Gly Leu
        35                  40                  45
Gln Asp Gly Gly Asp Asn Asp Asp Ala Ala Glu Ala Ser Ser Glu Pro
    50                  55                  60
Thr Glu Lys Ala Pro Ser Glu Glu Glu Leu His Gly Asp Gln Thr Asp
65                  70                  75                  80
Phe Gly Gln Gly Ser Gln Ser Pro Gln Lys Gln Glu Glu Gln Arg Gln
                85                  90                  95
His Leu His Leu Met Val Gln Leu Leu Arg Pro Gln Asp Asp Ile Arg
            100                 105                 110
Leu Ala Ala Gln Leu Glu Ala Pro Arg Pro Pro Arg Leu Arg Tyr Leu
        115                 120                 125
Leu Val Val Ser Thr Arg Glu Gly Glu Gly Leu Ser Gln Asp Glu Thr
    130                 135                 140
Val Leu Leu Gly Val Asp Phe Pro Asp Ser Ser Ser Pro Ser Cys Thr
145                 150                 155                 160
Leu Gly Leu Val Leu Pro Leu Trp Ser Asp Thr Gln Val Tyr Leu Asp
                165                 170                 175
Gly Asp Gly Gly Phe Ser Val Thr Ser Gly Gly Gln Ser Arg Ile Phe
            180                 185                 190
Lys Pro Ile Ser Ile Gln Thr Met Trp Ala Thr Leu Gln Val Leu His
        195                 200                 205
Gln Ala Cys Glu Ala Ala Leu Gly Ser Gly Leu Val Pro Gly Gly Ser
    210                 215                 220
Ala Leu Thr Trp Ala Ser His Tyr Gln Glu Arg Leu Asn Ser Glu Gln
225                 230                 235                 240
Ser Cys Leu Asn Glu Trp Thr Ala Met Ala Asp Leu Glu Ser Leu Arg
```

245 250 255
Pro Pro Ser Ala Glu Pro Gly Gly Ser Ser Glu Gln Glu Gln Met Glu
260 265 270
Gln Ala Ile Arg Ala Glu Leu Trp Lys Val Leu Asp Val Ser Asp Leu
275 280 285
Glu Ser Val Thr Ser Lys Glu Ile Arg Gln Ala Leu Glu Leu Arg Leu
290 295 300
Gly Leu Pro Leu Gln Gln Tyr Arg Asp Phe Ile Asp Asn Gln Met Leu
305 310 315 320
Leu Leu Val Ala Gln Arg Asp Arg Ala Ser Arg Ile Phe Pro His Leu
325 330 335
Tyr Leu Gly Ser Glu Trp Asn Ala Ala Asn Leu Glu Glu Leu Gln Arg
340 345 350
Asn Arg Val Thr His Ile Leu Asn Met Ala Arg Glu Ile Asp Asn Phe
355 360 365
Tyr Pro Glu Arg Phe Thr Tyr His Asn Val Arg Leu Trp Asp Glu Glu
370 375 380
Ser Ala Gln Leu Leu Pro His Trp Lys Glu Thr His Arg Phe Ile Glu
385 390 395 400
Ala Ala Arg Ala Gln Gly Thr His Val Leu Val His Cys Lys Met Gly
405 410 415
Val Ser Arg Ser Ala Ala Thr Val Leu Ala Tyr Ala Met Lys Gln Tyr
420 425 430
Glu Cys Ser Leu Glu Gln Ala Leu Arg His Val Gln Glu Leu Arg Pro
435 440 445
Ile Ala Arg Pro Asn Pro Gly Phe Leu Arg Gln Leu Gln Ile Tyr Gln
450 455 460
Gly Ile Leu Thr Ala Ser Arg Gln Ser His Val Trp Glu Gln Lys Val
465 470 475 480
Gly Gly Val Ser Pro Glu Glu His Pro Ala Pro Glu Val Ser Thr Pro
485 490 495
Phe Pro Leu Leu Pro Pro Glu Pro Glu Gly Gly Gly Glu Glu Lys Val
500 505 510
Val Gly Met Glu Glu Ser Gln Ala Ala Pro Lys Glu Glu Pro Gly Pro
515 520 525
Arg Pro Arg Ile Asn Leu Arg Gly Val Met Arg Ser Ile Ser Leu Leu
530 535 540
Glu Pro Ser Leu Glu Leu Glu Ser Thr Ser Glu Thr Ser Asp Met Pro
545 550 555 560
Glu Val Phe Ser Ser His Glu Ser Ser His Glu Glu Pro Leu Gln Pro
565 570 575
Phe Pro Gln Leu Ala Arg Thr Lys Gly Gly Gln Gln Val Asp Arg Gly
580 585 590
Pro Gln Pro Ala Leu Lys Ser Arg Gln Ser Val Val Thr Leu Gln Gly
595 600 605
Ser Ala Val Val Ala Asn Arg Thr Gln Ala Phe Gln Glu Gln Glu Gln
610 615 620
Gly Gln Gly Gln Gly Gln Gly Glu Pro Cys Ile Ser Ser Thr Pro Arg
625 630 635 640
Phe Arg Lys Val Val Arg Gln Ala Ser Val His Asp Ser Gly Glu Glu
645 650 655
Gly Glu Ala

<210> 6
<211> 408
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Ala Leu Val Thr Val Ser Arg Ser Pro Pro Gly Ser Gly Ala Ser
 1               5                  10                  15
Thr Pro Val Gly Pro Trp Asp Gln Ala Val Gln Arg Arg Ser Arg Leu
            20                  25                  30
Gln Arg Arg Gln Ser Phe Ala Val Leu Arg Gly Ala Val Leu Gly Leu
        35                  40                  45
Gln Asp Gly Gly Asp Asn Asp Asp Ala Ala Glu Ala Ser Ser Glu Pro
    50                  55                  60
Thr Glu Lys Ala Pro Ser Glu Glu Glu Leu His Gly Asp Gln Thr Asp
65                  70                  75                  80
Phe Gly Gln Gly Ser Gln Ser Pro Gln Lys Gln Glu Glu Gln Arg Gln
                85                  90                  95
His Leu His Leu Met Val Gln Leu Leu Arg Pro Gln Asp Asp Ile Arg
            100                 105                 110
Leu Ala Ala Gln Leu Glu Ala Pro Arg Pro Pro Arg Leu Arg Tyr Leu
        115                 120                 125
Leu Val Val Ser Thr Arg Glu Gly Glu Gly Leu Ser Gln Asp Glu Thr
    130                 135                 140
Val Leu Leu Gly Val Asp Phe Pro Asp Ser Ser Ser Pro Ser Cys Thr
145                 150                 155                 160
Leu Gly Leu Val Leu Pro Leu Trp Ser Asp Thr Gln Val Tyr Leu Asp
                165                 170                 175
Gly Asp Gly Gly Phe Ser Val Thr Ser Gly Gly Gln Ser Arg Ile Phe
            180                 185                 190
Lys Pro Ile Ser Ile Gln Thr Met Trp Ser Ser Glu Gln Glu Gln Met
        195                 200                 205
Glu Gln Ala Ile Arg Ala Glu Leu Trp Lys Val Leu Asp Val Ser Asp
    210                 215                 220
Leu Glu Ser Val Thr Ser Lys Glu Ile Arg Gln Ala Leu Glu Leu Arg
225                 230                 235                 240
Leu Gly Leu Pro Leu Gln Gln Tyr Arg Asp Phe Ile Asp Asn Gln Met
                245                 250                 255
Leu Leu Leu Val Ala Gln Arg Asp Arg Ala Ser Arg Ile Phe Pro His
            260                 265                 270
Leu Tyr Leu Gly Ser Glu Trp Asn Ala Ala Asn Leu Glu Glu Leu Gln
        275                 280                 285
Arg Asn Arg Val Thr His Ile Leu Asn Met Ala Arg Glu Ile Asp Asn
    290                 295                 300
Phe Tyr Pro Glu Arg Phe Thr Tyr His Asn Val Arg Leu Trp Asp Glu
305                 310                 315                 320
Glu Ser Ala Gln Leu Leu Pro His Trp Lys Glu Thr His Arg Phe Ile
                325                 330                 335
Glu Ala Ala Arg Ala Gln Gly Thr His Val Leu Val His Cys Lys Met
            340                 345                 350
Gly Val Ser Arg Ser Ala Ala Thr Val Leu Ala Tyr Ala Met Lys Gln
        355                 360                 365
Tyr Glu Cys Ser Leu Glu Gln Ala Leu Arg His Val Gln Glu Leu Arg
    370                 375                 380
Pro Ile Ala Arg Pro Asn Pro Gly Phe Leu Arg Gln Leu Gln Ile Tyr
385                 390                 395                 400
Gln Gly Ile Leu Thr Ala Arg Thr
                405
```

<210> 7
<211> 2704
<212> DNA
<213> Homo sapiens

<400> 7

cgtccttcct ggtcctgcgg gtccaggact gtccgcgggg ttgagggaag
gggccgtgcc 60
cggtgccagc ccaggtgctc gcggcctggc tccatggccc tggtcacagt
gagccgttcg 120
cccccgggca gcggcgcctc cacgcccgtg gggccctggg accaggcggt
ccagcgaagg 180
agtcgactcc agcgaaggca gagctttgcg gtgctccgtg gggctgtcct
gggactgcag 240
gatggagggg acaatgatga tgcagcagag gccagttctg agccaacaga
gaaggccccg 300
agtgaggagg agctccacgg ggaccagaca gacttcgggc aaggatccca
gagtccccag 360
aagcaggagg agcagaggca gcacctgcac ctcatggtac agctgctgag
gccgcaggat 420
gacatccgcc tggcagccca gctggaggca ccccggcctc cccggctccg
ctacctgctg 480
gtagtttcta cacgagaagg agaaggtctg agccaggatg agacggtcct
cctgggcgtg 540
gatttccctg acagcagctc ccccagctgc accctgggcc tggtcttgcc
cctctggagt 600
gacacccagg tgtacttaga tggagacggg ggcttcagcg tgacgtctgg
tgggcaaagc 660
cggatcttca agcccatctc catccagacc atgtgggcca cactccaggt
attgcaccaa 720
gcatgtgagg cagctctagg cagcggcctt gtaccgggtg gcagtgccct
cacctgggcc 780
agccactacc aggagagact gaactccgaa cagagctgcc tcaatgagtg
gacggctatg 840
gccgacctgg agtctctgcg gcctcccagc gccgagcctg gcgggtcctc
agaacaggag 900
cagatggagc aggcgatccg tgctgagctg tggaaagtgt tggatgtcag
tgacctggag 960
agtgtcactt ccaaagagat ccgccaggct ctggagctgc gcctggggct
ccccctccag 1020
cagtaccgtg acttcatcga caaccagatg ctgctgctgg tggcacagcg
ggaccgagcc 1080
tcccgcatct tccccccacct ctacctgggc tcagagtgga acgcagcaaa
cctggaggag 1140
ctgcagagga acagggtcac ccacatcttg aacatggccc gggagattga
caacttctac 1200
cctgagcgct tcacctacca caatgtgcgc ctctgggatg aggagtcggc
ccagctgctg 1260
ccgcactgga aggagacgca ccgcttcatt gaggctgcaa gagcacaggg
cacccacgtg 1320
ctggtccact gcaagatggg cgtcagccgc tcagcggcca cagtgctggc
ctatgccatg 1380
aagcagtacg aatgcagcct ggagcaggcc ctgcgccacg tgcaggagct
ccggcccatc 1440
gcccgcccca accctggctt cctgcgccag ctgcagatct accagggcat
cctgacggcc 1500
agaacctgag ggtggtgggg aggagaaggt tgtaggcatg gaagagagcc
aggcagcccc 1560
gaaagaagag cctgggccac ggccacgtat aaacctccga ggggtcatga
ggtccatcag 1620
tcttctggag ccctccttgg agctggagag cacctcagag accagtgaca
tgccagaggt 1680
cttctcttcc cacgagtctt cacatgaaga gcctctgcag cccttcccac
agcttgcaag 1740
gaccaaggga ggccagcagg tggacagggg gcctcagcct gccctgaagt
cccgccagtc 1800

agtggttacc ctccagggca gtgccgtggt ggccaaccgg acccaggcct
tccaggagca 1860
ggagcagggg caggggcagg ggcagggaga gccctgcatt tcctctacgc
ccaggttccg 1920
gaaggtggtg agacaggcca gcgtgcatga cagtggagag gagggcgagg
cctgagccct 1980
cacacatgcc cacgctcccc tgacactgaa gaggatccac aactccttgg
agaaacaccc 2040
tcacgtctgt tgccgcacac attcctctca gctccgcccc atacccgtca
ctacagcctc 2100
acctcccacc cctgtcacta cggcctcacc tcccaccccc gtcactacag
cctcacctcc 2160
tacagcctta agtcccaggc ccatgtctgc ctgtccaagg gctcaagact
ttctaactgg 2220
gatgtggtag agggactgaa ggtacctttg ggggcaacag caccctagtt
tcattctcaa 2280
ctctagccct gcacactcac ctgtggcacg gaatgaaaac agagcttccc
gtgcaaaaag 2340
ggtcacgcct cccacccccg ccccctccct gcacctcctg tcctctccca
gttcattcct 2400
ggaaccagcc aggccaggca accagtggcc cccaaaggca ggcaggatcc
tcaggcccca 2460
gccgcgggag gctggaaggg ctggcagatc gcttccctca tccacctcca
ccggtccagg 2520
tctttgctgc tgtccccaga cctcctgtga caccacgcca gatcacaggg
caccaggcca 2580
gagatagtct tcttttgtc ctttctggcc tctggctagt cagtttttca
tagccttaca 2640
gtatctggct ttgtactgag aaataaaaca cattttcata aaaaaaaaaa
aaaaaaaaaa 2700
aaaa
2704

<210> 8
<211> 312
<212> PRT
<213> Homo sapiens

<400> 8
Met Ala Leu Val Thr Val Ser Arg Ser Pro Pro Gly Ser Gly Ala Ser
1 5 10 15
Thr Pro Val Gly Pro Trp Asp Gln Ala Val Gln Arg Arg Ser Arg Leu
20 25 30
Gln Arg Arg Gln Ser Phe Ala Val Leu Arg Gly Ala Val Leu Gly Leu
35 40 45
Gln Asp Gly Gly Asp Asn Asp Asp Ala Ala Glu Ala Ser Ser Glu Pro
50 55 60
Thr Glu Lys Ala Pro Ser Glu Glu Glu Leu His Gly Asp Gln Thr Asp
65 70 75 80
Phe Gly Gln Gly Ser Gln Ser Pro Gln Lys Gln Glu Glu Gln Arg Gln
85 90 95
His Leu His Leu Met Val Gln Leu Leu Arg Pro Gln Asp Asp Ile Arg
100 105 110
Leu Ala Ala Gln Leu Glu Ala Pro Arg Pro Pro Arg Leu Arg Tyr Leu
115 120 125
Leu Val Val Ser Thr Arg Glu Gly Glu Gly Leu Ser Gln Asp Glu Thr
130 135 140
Val Leu Leu Gly Val Asp Phe Pro Asp Ser Ser Ser Pro Ser Cys Thr
145 150 155 160
Leu Gly Leu Val Leu Pro Leu Trp Ser Asp Thr Gln Val Tyr Leu Asp
165 170 175

```
Gly Asp Gly Gly Phe Ser Val Thr Ser Gly Gly Gln Ser Arg Ile Phe
            180                 185                 190
Lys Pro Ile Ser Ile Gln Thr Met Trp Ala Thr Leu Gln Val Leu His
        195                 200                 205
Gln Ala Cys Glu Ala Ala Leu Gly Ser Gly Leu Val Pro Gly Gly Ser
    210                 215                 220
Ala Leu Thr Trp Ala Ser His Tyr Gln Glu Arg Leu Asn Ser Glu Gln
225                 230                 235                 240
Ser Cys Leu Asn Glu Trp Thr Ala Met Ala Asp Leu Glu Ser Leu Arg
                245                 250                 255
Pro Pro Ser Ala Glu Pro Gly Gly Ser Ser Glu Gln Glu Gln Met Glu
            260                 265                 270
Gln Ala Ile Arg Ala Glu Leu Trp Lys Val Leu Glu Leu Glu Ser Thr
        275                 280                 285
Ser Glu Thr Ser Asp Met Pro Glu Val Phe Ser Ser His Glu Ser Ser
    290                 295                 300
His Glu Glu Pro Leu Gln Pro Phe
305                 310
```

<210> 9
<211> 524
<212> PRT
<213> Drosophila melanogaster

<400> 9

```
Met Ala Leu Val Thr Val Gln Arg Ser Pro Ser Val Ala Gly Ser Cys
 1               5                  10                  15
Ser Asn Ser Asp Gly Glu Ser Glu Asp Asp Glu Gly Asn Ser Lys Gly
            20                  25                  30
Asn Asp Arg Ser Glu Cys Phe Phe Ala Gly Lys Gly Thr Ala Leu Val
        35                  40                  45
Leu Ala Leu Lys Asp Ile Pro Pro Leu Thr Gln Ser Glu Arg Arg Leu
    50                  55                  60
Ser Thr Asp Ser Thr Arg Ser Ser Asn Ser Thr Gln Ser Asn Asn Ser
65                  70                  75                  80
Asp Ile Gln Leu His Leu Gln Ser Met Phe Tyr Leu Leu Gln Arg Glu
                85                  90                  95
Asp Thr Leu Lys Met Ala Val Lys Leu Glu Ser Gln Arg Ser Asn Arg
            100                 105                 110
Thr Arg Tyr Leu Val Ile Ala Ser Arg Ser Cys Cys Arg Ser Gly Thr
        115                 120                 125
Ser Asp Arg Arg Arg His Arg Ile Met Arg His His Ser Val Lys Val
    130                 135                 140
Gly Gly Ser Ala Gly Thr Lys Ser Ser Thr Ser Pro Ala Val Pro Thr
145                 150                 155                 160
Gln Arg Gln Leu Ser Val Glu Gln Thr Ala Thr Glu Ala Ser Ser Lys
                165                 170                 175
Cys Asp Lys Thr Ala Asp Lys Glu Asn Ala Thr Ala Ala Gly Asp Asn
            180                 185                 190
Lys Asn Thr Ser Gly Met Glu Glu Ser Cys Leu Leu Gly Ile Asp Cys
        195                 200                 205
Asn Glu Arg Thr Thr Ile Gly Leu Val Val Pro Ile Leu Ala Asp Thr
    210                 215                 220
Thr Ile His Leu Asp Gly Asp Gly Gly Phe Ser Val Lys Val Tyr Glu
225                 230                 235                 240
Lys Thr His Ile Phe Lys Pro Val Ser Val Gln Ala Met Trp Ser Ala
                245                 250                 255
Leu Gln Thr Leu His Lys Val Ser Lys Lys Ala Arg Glu Asn Asn Phe
            260                 265                 270
Tyr Ala Ser Gly Pro Ser His Asp Trp Leu Ser Ser Tyr Glu Arg Arg
```

```
        275                 280                 285
Ile Glu Ser Asp Gln Ser Cys Leu Asn Glu Trp Asn Ala Met Asp Ala
    290                 295                 300
Leu Glu Ser Arg Arg Pro Pro Ser Pro Asp Ala Ile Arg Asn Lys Pro
305                 310                 315                 320
Pro Glu Lys Glu Glu Thr Glu Ser Val Ile Lys Met Lys Leu Lys Ala
                325                 330                 335
Ile Met Met Ser Val Asp Leu Asp Glu Val Thr Ser Lys Tyr Ile Arg
            340                 345                 350
Gly Arg Leu Glu Glu Ile Leu Asp Met Asp Leu Gly Glu Tyr Lys Ser
        355                 360                 365
Phe Ile Asp Ala Glu Met Leu Val Ile Leu Gly Gln Met Asp Ala Pro
    370                 375                 380
Thr Lys Ile Phe Glu His Val Tyr Leu Gly Ser Glu Trp Asn Ala Ser
385                 390                 395                 400
Asn Leu Glu Glu Leu Gln Lys Asn Gly Val Arg His Ile Leu Asn Val
                405                 410                 415
Thr Arg Glu Ile Asp Asn Phe Phe Pro Gly Thr Phe Glu Tyr Phe Asn
            420                 425                 430
Val Arg Val Tyr Asp Asp Glu Lys Thr Asn Leu Leu Lys Tyr Trp Asp
        435                 440                 445
Asp Thr Phe Arg Tyr Ile Thr Arg Ala Lys Ala Glu Gly Ser Lys Val
    450                 455                 460
Leu Val His Cys Lys Met Gly Val Ser Arg Ser Ala Ser Val Val Ile
465                 470                 475                 480
Ala Tyr Ala Met Lys Ala Tyr Gln Trp Glu Phe Gln Gln Ala Leu Glu
                485                 490                 495
His Val Lys Lys Arg Arg Ser Cys Ile Lys Pro Asn Lys Asn Phe Leu
            500                 505                 510
Asn Gln Leu Glu Thr Tyr Ser Gly Met Leu Asp Ala
        515                 520
```

<210> 10
<211> 111
<212> PRT
<213> Homo sapiens

<400> 10

```
Met Ala Arg Glu Ile Asp Asn Phe Tyr Pro Glu Arg Phe Thr Tyr His
 1               5                  10                  15
Asn Val Arg Leu Trp Asp Glu Glu Ser Ala Gln Leu Leu Pro His Trp
            20                  25                  30
Lys Glu Thr His Arg Phe Ile Glu Ala Ala Arg Ala Gln Gly Thr His
        35                  40                  45
Val Leu Val His Cys Lys Met Gly Val Ser Arg Ser Ala Ala Thr Val
    50                  55                  60
Leu Ala Tyr Ala Met Lys Gln Tyr Glu Cys Ser Leu Glu Gln Ala Leu
65                  70                  75                  80
Arg His Val Gln Glu Leu Arg Pro Ile Ala Arg Pro Asn Pro Gly Phe
                85                  90                  95
Leu Arg Gln Leu Gln Ile Tyr Gln Gly Ile Leu Thr Ala Arg Thr
            100                 105                 110
```